DOCUMENT RESUME

ED 408 819 FL 024 508

AUTHOR	Butler, Priscilla, Ed.
TITLE	Issues in Alternative Assessment: The Japanese Perspective.
INSTITUTION	Kwansei Gaikun Univ., Nishinomiya (Japan). Language Center.
PUB DATE	97
NOTE	124p.
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Case Studies; *English (Second Language); Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Language Tests; Portfolios (Background Materials); Program Descriptions; Second Language Instruction; Self Evaluation (Individuals); *Student Evaluation; Theory Practice Relationship; Vocabulary Development; Writing Instruction
IDENTIFIERS	Japanese People; *Multiple Intelligences; Writing Samples

ABSTRACT

A collection of essays addresses issues in alternative assessment in English as-a-Second-Language (ESL) instruction. Chapter 1, "Toward a Definition of Alternative Assessment," (Priscilla Butler) gives an overview of the concept of alternative assessment and introduces some theoretical considerations. Chapter 2, "The Roots of Alternative Assessment: Epistemological Constructs and Current Challenges," (Butler) documents the socio-cultural, historical, and political backdrop for shifts in assessment practice, and highlights some current issues. Chapter 3, "Portfolio-Based Assessment," (Maricel Santos) presents the rationale for portfolio use in second language classrooms, provides examples, and outlines steps for classroom implementation of the method. Chapter 4, "Incorporating Alternative Assessment in the Teaching of EFL Writing: Responding to Multiple Drafts," (Shalle Leeming) contrasts the traditional product-oriented approach and an alternative process-oriented approach to student writing. Chapter 5, "Student Self-Assessment: A Viable Alternative," (John Chapman-Rienstra) examines various ways students can begin to assess their own learning and discusses the particular benefits for Japanese students. Chapter 6, "Interactive Vocabulary Quizzes: An Alternative Form of Assessment Based on Multiple Intelligence Theory," (Gina Keefer) provides rationale and practical ideas. Chapter 7, "What Do We Do Now? A Survey of Current Assessment Practices at KGU," (Butler) focuses on the situation at Kwansei Gaikun University (Japan). (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 408 819

Issues in Alternative Assessment

The Japanese Perspective

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Tanaoki Kanzaki
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Priscilla Butler, Editor

Language Center
Kwansei Gakuin University

FL024508

BEST COPY AVAILABLE

Issues in Alternative Assessment
The Japanese Perspective

Written by Language Center Instructors
Priscilla Butler, John Chapman-Rienstra,
Gina Keefer, Shalle Leeming, Maricel Santos

With grateful acknowledgement to
the members of the Alternative Assessment Research Committee
Priscilla Butler, Takaaki Kanzaki, Shinji Kimura, Toshio Koyama, Shalle Leeming,
Michael Lessard-Clouston, Gen Ohinata, Hiromi Ohtaka

This research project was made possible by Kwansei Gakuin University's Special Research Grant, funded in 1996.

© 1997, Language Center, Kwansei Gakuin University

Table of Contents

Preface iii
Overview iv

PART 1: THEORETICAL BACKGROUND

Chapter 1

Toward a Definition of Alternative Assessment 1
Priscilla Butler

Chapter 2

The Roots of Alternative Assessment: Epistemological Constructs and Current Challenges 10
Priscilla Butler

PART 2: CLASSROOM APPLICATIONS

Chapter 3

Portfolio-Based Assessment 23
Maricel Santos

Chapter 4

Incorporating Alternative Assessment in the Teaching of EFL Writing: Responding to Multiple Drafts 50
Shalle Leeming

Chapter 5

Student Self-Assessment: A Viable Alternative 55
John Chapman-Rienstra

Chapter 6

Interactive Vocabulary Quizzes: An Alternative Form of Assessment Based on Multiple Intelligence Theory 90
Gina Keefer

PART 3: CURRENT ASSESSMENT PRACTICES AT KGU

Chapter 7

What Do We Do Now? A Survey of Current Assessment Practices at KGU 95
Priscilla Butler

Preface

In some ways, the title of this book is somewhat ironic. After all, none of the chapter writers are Japanese, and none of us individually have been in Japan for more than six years. We chose this title deliberately, however, because we felt it was more reflective of what we were trying to do: research and think more deeply about the ways in which alternative assessments might work in Japanese classrooms.

In other ways, however, the literal meaning of the title is at the same time apt. Those of us teaching in the Language Center at Kwansei Gakuin University do not often have contact with teachers in other programs or departments. By working with the other Japanese and foreign faculty members on the research committee which was formed after we proposed this topic, we learned much more about the context in which most of us had been teaching for quite some time. In some ways, it took a research process like this to learn more about the everyday: including what other English teachers at our own campus are thinking and doing about assessing their students' learning. The discussions with our Japanese colleagues were particularly formative in helping us to better understand our task.

The idea to research and write about alternative assessment itself grew out of discussions with other Language Center instructors who were keenly interested in the new models of assessment and who wanted to learn more about them. When we decided to apply for university funding for this research project, we believed that our own thinking would benefit from working with others who were also framing similar questions about their assessment practice. With the additional support of faculty in other programs, we began to explore what it means to revision the way we think about assessment.

As we began to write in order to put across what we had learned, we were faced with an additional challenge: trying to decide who are audience might be. Though we could be sure that some Kwansei Gakuin English teachers would read our work, we could not be sure of their individual academic and professional backgrounds. Because many college and university English teachers in Japan are experts in Literature or Linguistics rather than TESOL or Applied Linguistics, we envisioned a broadly based audience with varying degrees of background knowledge about ESL/EFL pedagogy and theory.

Once we decided to list our publication on international databases, we multiplied both the numbers and types of readers who might eventually read our work. Considering that some of these readers might not be familiar with current TESOL practice, we tried to write in ways that would at once give enough background to those new to the field while not supplying too much detail so as to bore those already familiar with the basic concepts. Additionally, we tried to keep in mind that many of our readers might speak English as a second language and might be most familiar with academic prose. These expectations also influenced our writing.

Of course, we faced only what every writer faces: the dilemma of understanding audience before writing. We hope we have gauged the interests and needs of our readers correctly, and that this volume serves as a useful addition to your own library, wherever you may be.

Overview

The seven chapters in this volume explore alternative assessment theory and practice from very diverse angles. Chapter 1 provides a broad overview of alternative assessment in order to define what people actually mean when they use the term. It additionally functions to introduce some of the theoretical constructs underlying alternative assessment practice.

Chapter 2 examines theoretical issues from a more contextualized angle, documenting the socio-cultural, historical, and political backdrop for shifts in assessment practice while also highlighting some of the major issues surrounding current alternative assessment practice.

The next four chapters focus on practical applications of alternative assessment to the classroom, specifically the Japanese college or university English class. In chapter 3, Maricel Santos outlines the rationale for doing portfolios in a language classroom, providing extensive examples from the classes she has taught. The second half of this chapter includes a step by step outline of how to implement the portfolio process in a language class. Ms. Santos includes detailed answers to some of the most common questions teachers may have about using portfolios along with eight original handouts teachers may use with their students if they wish to begin working with portfolios in their own classrooms.

In chapter 4, Shalle Leeming looks at ways of responding to student writing: contrasting a more traditional product-oriented approach with an alternative process-oriented one. In addition to outlining some of the reasons teachers may wish to adopt a process oriented approach, Ms. Leeming gives useful hints for responding to students writing, both on a content level as well as a grammatical level. She documents common techniques used to respond to multiple drafts of student work.

In chapter 5, John Chapman-Rienstra examines various ways that students can begin to assess their own learning. First, he explicates how he himself came to believe that student self-assessment would especially benefit Japanese students. In the remainder of his paper, he provides detailed descriptions of some of the major activities he does in his classes which involve student self assessment, including oral vocabulary quizzes, reading discussion groups, and debates. Eighteen pages of class handouts accompany this article.

Chapter 6 focuses on vocabulary quizzes that students can take interactively. Gina Keefer explains how the concept of multiple intelligences can be coupled with this class assessment tool to provide an alternative format for students to measure their learning.

The final chapter in this book moves back into the community of the university, focusing on the current assessment practices of Kwansei Gakuin University English teachers. Results from a university-wide survey of English teachers are presented and then followed by a discussion of some of the main issues concerning Japanese English education at the university level.

Part 1

Theoretical Background

Chapter 1

Toward a Definition of Alternative Assessment

PRISCILLA BUTLER

> I was picking up my fourth-grade daughter after school one afternoon when, as she jumped into the car, she exclaimed, "Mom, I've never in my life had none of the above!" I thought for a moment and then realized just what she was talking about--a multiple choice test. (Huerta-Macias, 1995, p. 8)

> This kind of evaluation is much harder than giving a multiple-choice test, but if giving multiple-choice tests were central to the assessment of human growth and development, we parents would administer them regularly at home. Instead we lie in bed at night asking ourselves and our mates exactly the kinds of questions I've suggested . . . and then--together with our kids--we decide, and we roll up our sleeves and work together to accomplish what we deem important. We should operate the same way in schools. (Noddings, 1992, p. 180)

The most common definitions of alternative assessment are stated in negative terms, that is alternative assessments are commonly defined in terms of what they are not. They are not multiple-choice, machine scorable, paper and pencil tests. This definition gives us at least some notion of the parameters of the term, but, in order to better understand what we do when we alternatively assess, I think it important that we share similar understandings of what it means to alternatively assess.

A SHORT HISTORY OF LANGUAGE TESTING

Looking at the historical roots of language testing is one way to see how this most recent innovation has developed. In his historical overview of language testing, Spolsky (1995) defines three eras of testing: traditional, modern, and post-modern. For those familiar with his prior work on language testing, Spolsky had previously used different titles for these three periods: traditional or pre-scientific, psychometric-structuralist or scientific, and psycholinguistic-sociolinguistic (Spolsky, 1977). In this initial section, I will use these words interchangeably to help explicate the divisions between the three periods of testing.

Spolsky dates the traditional or pre-scientific period of testing back 2000 years to the Chinese civil service examinations. These examinations were first offered in China as a more objective means of measuring who would be selected to become civil servants than the previous means of selection through patronage. In the late sixteenth century, the first Western scholarly account of these exams was given by a French Jesuit explorer and missionary, and the traditional type of exam was adopted in French

Christian schools in the eighteenth century: eventually spreading to other parts of Europe, including England as well as the United States by the nineteenth century (Spolsky, 1995, pp. 16-18). These traditional exams were primarily written essay which were graded anonymously by teachers. The exam system assumed that students had to memorize substantial amounts of information and demonstrate knowledge of that information in writing. The current practice of giving essay exams is one illustration of this traditional approach to testing.

The second phase of testing, modern or psychometric, amazingly enough only has a history of about one hundred years. Perhaps the most well known current example of a psychometric language test is the TOEFL (Test of English as a Foreign Language). The origins of psychometric testing are found in the field of psychology (hence the "psycho" at the beginning of the name). Prior to the development of a more systematic, reliable measure to test students' learning, traditional tests had already come under fire for the inconsistency of the scores assigned to them depending on who graded them at what time. A British professor of statistics, F. Y. Edgeworth published his speculations in a series of papers first appearing in 1888 in which he suggested that examinations were too fraught with error: even as high as 13 percent in one study, enough to make or break a passing score on the crucial exams (Spolsky, 1990, pp. 23-25). He suggested that examiners must be meticulous in keeping these realities in mind and equally careful to reduce the possibility of extreme discrepancies between raters.

While psychometric tests did not arise in response to Edgeworth's work, they certainly served to address several of the issues he had raised. The modern test arose out of the work of a French doctor, A. Binet, who sought a more useful means of determining "which children were retarded and needed special education" (Spolsky, 1990, p. 27). Binet sought a way to measure intelligence, and his work eventually grow into early forms of I.Q. tests. As this work spread to the continent (the U.S.), it began to have effects on the field of testing itself. Backed by the claims that these new psychological measures had established scientific objectivity, not to mention reliability and validity, psychometric testing theory became central to the idea of testing itself, ironically much more so in the U.S. than in Europe itself where the initial concepts of psychometric testing derived. As Spolsky notes, "By 1920, most state universities were offering courses in educational measurement, and by 1923 half the business of the Teachers College Bureau of Publications was in test and scales" (1995, p. 37; based on Joncich 1968, p. 389). Various social and political influences pole-vaulted psychometric tests into a position of prominence in American education--a position that would not be seriously challenged for some years to come.

The challenge to the "objective" modern tests did finally come, in the language field as early as 1961 when calls were sounded to develop language tests that would assess language integratively rather than discretely (Spolsky, 1995, p. 193). Bachman (1990: 300) also documents these calls for changes to the "objective" tests, quoting a passage from J. B. Carroll (1961). "I recommend tests in which there is less attention paid to specific structure-points or lexicon than to the total communicative effect of an utterance . . . . Indeed, this 'integrative' approach has several advantages over the 'discrete structure-point' approach" (p. 37). This call for a more authentic language test would be repeated in later years, especially as more communicative approaches gained precedence in theoretical and pedagogical constructs. This post-modern testing movement saw objective tests as being less able to really capture or say very much about the underlying abilities of language users. Criticisms that such objective tests may unfairly disadvantage certain types of learners (Garcia & Pearson, 1991; Schwartz, J., 1979) provided a further impetus for calling for a new approach to testing--neither a completely traditional, nor completely psychometric one.

Especially for young teachers who may only be acquainted with one of more of the testing strands described here, it might be harder to challenge existing patterns if it is believed that there is and has always only been one accepted method of testing. What Spolsky's work shows is that the history of formal testing is quite short, "objective,"

psychometric tests have only been around for about a hundred years or so, and that the division of language testing development into three periods might be more heuristically described as "an unresolved (and fundamentally unresolvable) tension between competing sets of forces" (p. 354). In other words, testing movements represent philosophical and practical movements.

A PHILOSOPHICAL RATIONALE FOR ALTERNATIVE ASSESSMENT

As Ryan and Miyasaka (1995), two educational researchers, note, alternative assessment and other systems of testing are not merely different activities; they also represent "basic beliefs about what students should learn and how they can best be taught" (p. 1). Other theorists have described what they believe are the divisions between more traditional modes of education and that promoted by supporters of alternative assessment and alternative education.

> The culture of testing extends far beyond the specifics of item design and test construction. It involves a network of activities that extend to the conduct of testing, the forms of reporting data, and the uses to which those data are put. Chief among these practices are the emphasis on relative ranking above actual accomplishment; the privileging of easily quantifiable, rather than messy and complex, displays of skills and knowledge; the assumption that individual performances, rather than collaborative forms of cognition, are the most powerful indicators of educational progress; and an image of evaluating educational progress as a matter of scientific measurement. (p. 43)

In other words, different ways of envisioning testing/assessment reflect different visions of view the educational process itself.

Proponents of alternative assessment view change in assessment as a change in the way curricula are envisioned. This is a movement away from thinking of the student as an object into which the maximum amount of information must be crammed toward thinking of the student as a knower who interacts with other knowers, which may include other students as well as teachers and those outside the school. Learning is viewed as an active process of thinking through complex ideas, the fruits of which the student can later apply in diverse settings. In this view, the student is an active part of the learning process, the teacher is not a repository of knowledge but rather a facilitator of learning, and classmates and outside of school experiences can be equal if not more important sources of learning. Ewell (1991) calls this philosophical repositioning as a movement "from an additive 'production process' view of education to one that recognizes complex, often nonlinear paths of development" (p. 87). The production model assumes that students learn information in small, precise units that progress gradually from simple to less simple. The alternative model suggests that students learn in multiple ways and that learning doesn't always occur in a straight line.[1]

DEFINING TERMS

Before looking more extensively at beliefs or practices of alternative assessment, it would be helpful to examine the words themselves so that a confusion does not exist about what practices are actually being described. A survey of the professional

[1] For a more extended discussion of the theoretical background, see chapter two of this manual, or the descriptions of cognitive learning theory described in Calfee, 1994 and Herman, Aschbacher, and Winters, 1992: pp. 14-17.

literature reveals that alternative assessment is often used as an umbrella term for any non-traditional assessment practice. Not everyone is satisfied with the descriptive power of this term (see Balliro, 1993, for example), but it has become quite widespread, perhaps because of its very inclusiveness and lack of detail. As Niyogi (1995) observes, "The descriptive terms . . . attempt to distinguish these alternative forms of assessments from standardized, primarily multiple-choice, paper-and-pencil tests. The terms authentic and performance are often used interchangeably to describe assessments that engage students in more 'hands-on' type activities and require them to create a product or construct a response" (p. 12).

Niyogi goes on to define another term, "naturalistic assessment" as being "rooted in the natural setting of the classroom and involving observation of student performance and behavior in a less structured, more informal context" (p. 12). While authentic and performance assessments generally try to mirror tasks that students would typically do in real life outside of the classroom, naturalistic assessment and some other forms of alternative assessment, while calling on students abilities to use skills they will need in their lives outside of school, do not necessarily seek to imitate actual activities students might encounter in future professional or personal contexts.

Other terms that have been used include "classroom-referenced assessment" (Genesee, 1994), referring to those activities which are based squarely in the classroom and which become an ongoing part of classroom life, providing useful feedback about instruction as well as about students. Herman, Aschbacher, and Winters (1992) provide a useful summary of these multiple definitions and what they actually mean.

> We use these terms [alternative assessment, authentic assessment, and performance-based assessment] synonymously to mean variants of performance assessments that require students to generate rather than choose a response. Performance assessment by any name requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems. Exhibitions, investigations, demonstrations, written or oral responses, journals, and portfolios are examples of the assessment alternatives we think of when we use the term "alternative assessment." (p. 2)

In understanding the breadth of activities that may be encapsulated under the large umbrella terms, "alternative assessment," it would be useful to look at some of the other lists that writers on education have generated to exemplify the range of assessment practices. Genesee (1994, p. 3) lists "portfolios, interactive diaries, student conferences, and anecdotal classroom observation" as some useful alternative assessment practices. Huerta-Macias (1995, p. 9) cites "checklists of student behaviors or products, journals, reading logs, videos of role plays, audiotapes of discussions, self-evaluation questionnaires, work samples, and teacher observations or anecdotal records" as being examples of some alternative assessment procedures. Tillyer and Sokolik (1993) discuss student-oriented projects as examples of alternatives, while Ryan (1995, p. 5) adds another alternative: scoring rubrics or rating scales.

It is to this last category that much attention has recently been given. Two excellent guides to understanding rubrics and how they fit into alternative assessment practice are Herman, Aschbacher, and Winters (1992) and Marzano, Pickering, and McTighe (1993). Also, Mentkowski and Loacker (1985) and Courts and McInerney (1993) provide detailed case studies of how general alternative assessments were implemented along with curriculum innovations at colleges in the U.S., while O'Malley and Valdez Pierce (1996) provide multiple, practical examples of different types of assessment that would be suitable at levels from elementary through college.

Though it would take too much space to define each of the myriad of alternative assessment procedures in detail here, much less give examples of how they operate in practice, I assume that many of these assessments are already familiar to a substantial

number of language teachers and that those who would like to read more about them can turn to any of the numerous resources on the subject. Without a deeper understanding of the basic tenets of alternative assessment, however, any more detailed discussion would have questionable value.

BASIC BELIEFS

One of the most central tenets in the practical application of alternative assessments is that each student should be assessed in a variety of ways and should have multiple opportunities to demonstrate learning in various modes (Herman, Aschbacher, & Winters, 1992; O'Malley & Valdez Pierce, 1996). As Niyogi (1995) remarks, "It is critical to remember that an assessment program by definition should employ some *breadth and variety* of strategies and procedures for observing, collecting, and evaluating student work and student learning" (p. 12, emphasis in original). As research on learning styles (Faltis, 1996; Gardner, 1983; Reid, 1995) and cognitive learning theory (Calfee, 1994; Herman, Aschbacher, & Winters, 1992) documents, students have differing ways of receiving, processing, and demonstrating understanding of ideas. Alternative assessment practitioners believe that if students are asked to demonstrate their understanding in only one mode, they will be unfairly disadvantaging those students who may not process well in that mode.

An example would serve well in making this idea more concrete. If there is a student who learns best when he or she is able to talk through ideas with others and who processes information best when he hears it rather than sees it, then to give this student *only* a written essay exam might not allow him or her the opportunity to really show what he or she knows. Likewise, for visual, analytical learners, such an essay exam would be comparatively easy, and they would be more likely to perform well on them and appear to have learned the concepts more deeply.

One assumption of alternative assessment practice, then, is that to rely too heavily on one and only one instance of a particular type of assessment unfairly disadvantages some students while unfairly advantaging others. Proponents would contend that to more fully capture the complexity of a student's learning, multiple ways of viewing a student's performance on many occasions over an extended period of time are crucial. Such a system, for example, would more strongly support evaluating students on more than just one measure, like a final exam. As Marzano, Pickering, and McTighe (1993) and Ryan and Miyasaka (1995) have noted, though, alternative assessment practice *does not exclude "objective" or traditional tests* as being one potential source of information about a student's learning; rather, it posits that such *"objective" tests should not be the only or primary way* of understanding the complexity of a student's learning.

A second central assumption is that learning assessments, including those done by teachers, classmates, or the students themselves, should provide feedback to learners so that they have an opportunity to learn from their assessments. In other words, *assessment should serve the learning process rather than simply evaluating it.* In this way, assessment becomes a tool for learning and are not simply initiated at the stage that a student is ready to complete a product. O'Malley and Valdez Pierce (1996) have noted that one of the qualities of alternative assessment is "that the criteria are made public and known in advance" (p. 5, based on Aschbacher, 1991). By letting students know what is expected of them in clear terms before students begin work, and by providing feedback at various stages of the learning process, students benefit from ongoing assessment rather than summative or end-stage assessment.

This second tenet, then, feeds into the first in that for assessment to be applied in a constructive way to future classroom efforts, it should be part of normal classroom procedures: ongoing and repeated on multiple occasions. Ana Huerta-Macias summarizes some of the features of alternative assessment in a very similar way:

> Alternative assessment, most importantly, provides alternatives to traditional testing in that it (a) does not intrude on regular classroom activities; (b) reflects the curriculum that is actually being implemented in the classroom; (c) provides information on the strengths and weaknesses of each individual student; (d) provides multiple indices that can be used to gauge student progress; and (e) is more multiculturally sensitive and free of norm, linguistic, and cultural biases found in traditional testing. (p. 9)

This focus on the classroom as the major arena of assessment links well with another important aspect of assessment that is sometimes forgotten.

UNGRADED ASSESSMENTS

Many teachers associate assessment with only those practices or evaluation systems that will directly contribute to student grades; yet, ungraded assessments can be equally beneficial to the teaching and learning processes. As Donald Freeman (1996) has argued, "Teaching is a complex, messy business of knowing what to do in the classroom" (p. 170). Teachers constantly make decisions about what to do in response to the multiple factors that impinge on the lesson, crucially their observations of what students appear to be doing and how this feedback affects what teachers choose to do in response. Much of this goes on at an unconscious level. Many teachers rarely ask students directly about how they are learning in a classroom, much less how a particular mode of teaching seems to suit their learning processes, unless it is in the form of a final course evaluation that is clearly too late to affect the students who actually complete them. Of course, informal observations of what students *appear* to be doing or learning can be very useful; however, there are also other, more direct means of getting feedback on how students are experiencing the learning process.

In *Classroom assessment techniques: A handbook for college teachers* (1993), Angelo and Cross propose that asking, in a variety of ways, for more extensive, ongoing, specified feedback throughout the class can greatly increase a teacher's awareness of how what is happening is actually contributing to student's learning. Of the fifty different techniques they describe, perhaps the most well-known is the "One-Minute Paper." They describe a very simple yet pedagogically valuable task of asking students to reflect on their learning. In turn, teachers themselves can gather more concrete evidence of students' successes and problems than they may be able to gather from observation and intuition. The one-minute paper is done anonymously by each student in the class, and, as it's name suggests, is limited to a very short period of time, though not necessarily just one minute. At the end of class, the teacher would ask the students to answer two questions on a half-sheet of paper: "What was the most important thing you learned during this class?" and "What important question remains unanswered?" (pp. 148). This is just one idea of how teachers can use ongoing assessment to benefit their understanding of what is needed in teaching and to better understand what students need.

CONCLUSION

The body of literature on alternative assessment is extensive, especially when one looks outside of just one specific field and to a more general field like education. What has been written here is, and can only be, a very brief synopsis of some of the issues I believe are key to any discussion about alternative assessment. Chapter two provides a more in-depth, theoretical analysis with a focus on the contexts surrounding the movement toward alternative assessment. Before continuing to describe the influences that converged to create such a movement, however, I think it only fitting to

take a moment to recognize the very real contextual constraints that prohibit the adoption of alternative assessment practices in many Japanese educational institutions.

Class size and teaching load are just two of the most weighty barriers to adopting such assessments. Clearly, teachers can only do what is humanly possible. Calls for teachers to understand, individually observe, respond to, and give feedback on multiple classroom performances and assessments, then, are not only calls for changes in the way we think about assessment but also changes in the way we think about curricula, teaching practice, and educational institutions themselves. In explicating some of the possibilities of alternative assessment, I do not wish to deny the challenges that such practices may have in the Japanese context. Nor do I wish to ignore the very real cultural concerns that may impinge on these discussions. The echoes of these and other concerns will be returned to in greater depth, especially in chapters two and seven.

REFERENCES

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education, 4* (4), 275-288.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Balliro, L. (1993). What kind of alternative? Examining alternative assessment. *TESOL Quarterly, 27* (3), 558-561.

Calfee, R. (1994). *Implications of cognitive psychology for authentic assessment and instruction* (Tech. Rep. No. 52). Berkeley: University of California, National Center for the Study of Writing.

Carroll, J. B. (1961). Fundamental consideration in testing for English language proficiency of foreign students. In *Testing the English Proficiency of Foreign Students* (pp. 30-40). Washington, DC: Center for Applied Linguistics.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Higher Education, 17*, 75-125.

Faltis, C. (Ed.). (1996). Learning styles and strategies [Special issue]. *TESOL Journal, 6* (1).

Freeman, D. (1995). Redefining the relationship between research and what teachers know. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom* (pp. 134-181). Cambridge: Cambridge University Press.

Garcia, G. E., & Pearson, P. D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education* (pp. 337-391. Washington, DC: American Education Research Association.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5* (1), 8-11.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model.* Alexandria, VA: Association for Supervision and Curriculum Development.

Niyogi, N. S. (1995). *The intersection of instruction and assessment: The classroom.* (Report No. TM 024 199). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 388-724)

Nodding, N. (1992). *The challenge to care in schools: An alternative appraoch to education.* New York: Teachers College Press.

O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers.* Reading, MA: Addison-Wesley.

Reid, J. (Ed.). (1995). *Learning styles in ESL/EFL classrooms.* Boston: Heinle & Heinle.

Ryan, J. M., & Miyasaka, J. R. (1995). Current practices in testing and assessment: What is driving the changes? *NAASP Bulletin, 79* (573), 1-10.

Schwartz, J., L. (1979). *What do we know about teaching and learning in urban schools? Volume 3: Assessment that respects complexity in individuals and programs.* (Report No. UD020353). Washington, D. C.: National Institute of Education. (ERIC Document Reproduction Service No. ED 185 159)

Spolsky, B. (1995). *Measured words.* Oxford: Oxford University Press.

Chapter 2

The Roots of Alternative Assessment: Epistemological Constructs and Current Challenges

PRISCILLA BUTLER

> We have only to speak of an object to think that we are being objective. But, because we chose it in the first place, the object reveals more about us than we do about it. (Gaston Bachelard, in McGrane, 1989, p. ix)
>
> We see the world the way we do not because that is the way it is but because we have these ways of seeing. (Ludwig Wittgenstein, in McGrane, 1989, p. ix)

The past few years have witnessed a preponderance of published documents relating to alternative assessment of second and foreign language education, including books (Brindley, G., 1995; Genesse, 1996; Brown and Yamashita (1995), Hancock, C., 1994; Hill, C., & Parry, K., 1994; Homlar Fradd, S., Larrinanaga McGee, P., & Wilen, D. K., 1994; Law, B., & Eckes, M., 1995; O'Malley, J. M., & Valdez Pierce, L., 1996), entire or partial professional journals (Faltis, C, 1995, 1996a), and language teaching conferences (JALT, 1995), with their resulting published proceedings (van Troyer, Cornwell, Morikawa, 1995). In addition to this partial sampling of recent publications ranging from theoretical explorations, practical "how-to's" and historical "This is how we did it" pieces, a new spectrum of testing books has also been released, notably Bachman (1990), Brown (1996), and Alderson (1995). Many of the materials printed seem to have the objective of demystifying the testing/assessment puzzle, as in Alderson and Clapham (1995). Additionally, a survey of current scholarship suggests that language educators and theorists alike are re-examining what it means to assess students' progress.

Of course, this partial sampling of recent publications in the fields of Applied Linguistics/TESOL/SLA is scant in comparison with parallel publications in other education fields. Because alternative assessment traces its roots to the more general field of education, however, examining alternative assessment only within the confines of TESOL would be a mistake. A broader, more far-ranging perspective is needed to gain a clearer understanding of both the motivations behind alternative assessments movements and the relevance of such movements to current practice.

WHY NOW?

Those following the changes in education commentary during the last several years might be led to believe that alternative assessment and its many related terms is a recent development. However, the term itself can be traced back to articles as early as 1974 (Popham), while according to Rudner & Boston (1994) a related term like

"performance tests" has been used as a descriptor "since the birth of the ERIC system in 1966" (p. 4).[1]

As Spolsky (1995) has noted in his analytical treatment of the history of language testing, testing practice is slow to catch up with theoretical findings or pedagogical arguments for change. In his probing examination of the two major institutionalized English tests in the world--the TOEFL, an American test administered by ETS, the Educational Testing Service, and the ELTS (now the IELTS, International English Language Testing Service) as well as other proficiency tests administered by the University of Cambridge Local Examinations Syndicate (UCLES) in the United Kingdom--Spolsky posits that TOEFL administrators had access to theories of communicative approaches and the concept of communicative competence some twenty years before it began to make substantial changes to its testing practice by adding an optional spoken portion (the TSE, Test of Spoken English) and written test (the TWE, Test of Written English) to the exams. Changes currently being considered by ETS under the auspices of TOEFL 2000, a long-term research program, include making the test more interactive. The viability of administrating the test via computer and giving test-takers interactive tasks and immediate feedback on performance forms a major portion of the TOEFL 2000 research agenda. Spolsky suggests, however, that these changes were put off for a number of years because of a multitude of practical, ideological and political reasons, not the least of which is "the relative invulnerability of institutions to change" (p. 341).

In a similar path of inertia, UCLES took a substantial period of time to modify its own tests in response to the huge body of knowledge gained about "objective" testing. Spolsky documents the research process that showed how many of the qualities associated with psychometric testing, especially reliability and objectivity, were virtually ignored until an UCLES sponsored research report in 1989 outlined major flaws in reliability, despite the fact that criticisms of tests which had no demonstrated reliability surfaced as early as 1888 in England (p. 338), crucial aspects of psychometric test theory were introduced in 1905 (p. 27) and psychometric testing was relatively established and widely known by the 1920s (p. 37). The current revisions of IELTS finally take psychometric testing theory into account.

These two historical examples of English language testing change may help illustrate why calls for testing reformation in broader educational circles away from psychometric, multiple-choice type tests and toward alternative assessment appeared as early as the 1970s yet are now being taken seriously on a broad scale in the 1990s. Institutional change takes time.

Socio-Cultural Factors

Of course, things are never as simple as that. The alternative assessment movement, at least as it manifests itself in the U.S., also has broader socio-political motivations to change now. Multiculturalism and its attendant awareness of the diversity of student needs and levels of academic preparation is a major focus of U.S. education. The United States has always been an immigrant society, of course, but more and more public documents recognize the need to help schools accommodate changing populations. The 1990 United States census documented a staggering increase in minority populations, the biggest decade change in the entire 20th century, with nearly one in four Americans claiming African, Asian, Hispanic, or American Indian ancestry, compared with one in five the decade before (New York Times, 1990). If such demographic trends continue, by the year 2000 the numerical majority of the population of the United States could be comprised of ethnic and racial minority

[1] ERIC--Educational Resources Information Center--is the single largest English-language reference tool for educational research in the world.

groups, many of whom do not speak English as a first language. Research has documented that multiple-choice, recall type tests can unfairly disadvantage such learners (Garcia & Pearson, 1991; Schwartz, J., 1979).

The Influence of Learning Theory

Research into learning styles (Faltis, 1996b; Gardner, 1983; Reid, 1995) and theories of cognitive development (Calfee, 1994; Herman, Aschbacher, & Winters, 1992) suggest that even learners who do not fall into traditional minority groups also represent diverse/learning approaches and needs. As Knefelkamp (1992) argues, a recognition of diversity, especially in education, is essential to maintaining the conditions of peaceful interaction, as a class, as a neighborhood, as a nation, and as a planet.

Early associations of alternative assessments primarily with "disadvantaged" populations, including remedial first language learners, (Hill & Parry, 1994), do not take into account the diversity that individuals even of homogeneous race and language bring to learning situations. Mark Hicks (1995) argues that, "A universal realization of every human is the notion that each of us holds the values of more than one culture simultaneously. If one were to think of culture as codes of language, knowledge, skills, values and beliefs, then this becomes possible" (p. 2). He offers an alternative vision of diversity to encompass the individual as well: including characteristics such as gender, socio-economic class, religious tradition, sexual orientation, age, and race.

Both Knefelkamp and Hicks propose that the danger that exists in homogenizing entire groups by the categories (or lack of them) that we apply is that individuals may be denied parts of their experience, especially if they fall into traditionally labeled minority groups, while those in dominant groups are asked to believe in a uniform reality that may be no more heuristic in describing their experiences. For both groups, however, an accompanying danger is the continual marking of barriers between groups. If one group is thought to be homogeneous and other groups are different, it is easier to require others to fit a dominant mold just as it is easier to view others as deficient or even un-human if they do not fit a dominant group conception of how to act, speak, think, or learn.

In this view, alternative assessments are not just a tool for recognizing and accommodating differing styles of those from minority groups but also for recognizing the differing styles of those within single groups. Alternative assessments represent a new way of perceiving *human* learning rather than just *minority* learning. Of course, this idea too is not so new in many ways. In 1934, John Dewey put forward a model of education that sought to acknowledge differences in how individual students may learn. Alternative approaches to education have been around for quite some time, although the research on cognition that provides a more detailed frame for pedagogical practice is a more recent development. Now more than ever, at least in the United States, educators have recognized a need for a system of assessment practice that includes rather than excludes and martials the possibilities that all students bring to the classroom rather than seeking to rank order students against a norm unfairly suited to the populations as well as individual experiences from which they come.

APPLICABILITY TO THE JAPANESE CONTEXT

This view of individual differences may seem curiously American. Japanese educators may wonder how such an argument fits their own educational context, especially since, at least on a surface level, Japan is still a relatively homogeneous country. Also, such dramatically divergent views of knowledge, personality and character growth, and the learning process itself when applied to the Japanese

educational context might seem culturally insensitive at the least, if not downright culturally imperialistic, as the act of English language teaching has sometimes been charged (Pennycook, 1989; Philipson, 1992).

I provide this emerging view of what it means to learn and what it means for learners to learn in certain ways only to explicate the social, political, and theoretical constructs that provide backbone to the alternative assessment movement. At the same time, I recognize not only the marked difference in the theoretical models of education but also the differences in form. Generally larger class sizes, more structured classroom practices, and the more centralized system of educational control coming from the central government in the form of Ministry of Education guidelines for educational practice all mark distinct differences between North American and especially U.S. classrooms and those in Japan. When cultural differences are weighed into the balance, it becomes all the clearer that I do not intend this discussion as a prescription for how Japanese language educators, much less Japanese educators in general, should or should not teach in their classrooms.

Instead, I point out the rationale behind alternative assessment practice to provide food for thought. That alternative assessment is not just about minority learning or about the different ways that language minority learners may encounter learner situations but about the ways in which all individuals may differ on a variety of traits that affect their learning is an important point for consideration. Much talk has been devoted to the perceived need to internationalize Japanese educational curricula. If internationalization is viewed in the sense that greater understanding is developed for all those who differ from one's self, even if the person is one who sits in the next seat in class, then it would seem that discussions about formation of identity and learning beliefs would be a critical part of curriculum innovation. It is true that Japan is still a relatively homogeneous country compared with the United States; however, if diversity is viewed not only as cultural diversity in the sense that most of us are accustomed to thinking of it (based on race and/or national origin or language), then Japan too would be defined as a country whose educational system serves diverse student populations.

As I continue the discussion of diversity and its relationship to cognitive learning theory and, consequently, to alternative assessment models, I wish to keep this definition of diversity in mind: that each individual embodies diversity in her or his approach to learning because of the different characteristics she or he brings to school. While I do not wish to impose these definitions or theories on Japanese educators or even on educators in other countries or those in the United States who hold differing theoretical positions on the nature of learning, I offer this background as an important conceptual girder for the constructs of diversity and alternative assessment as they relate to theories of learning discussed in greater depth later in this paper.

HISTORICAL BACKGROUND

It may seem strange to discuss diversity in the context of language education. Issues of multicultural education that are so newly prominent in mainstream education publications are currently given less focus in TESOL-related fields. After all, teachers of English as a second language have always had to recognize the multicultural realities of their classrooms, although some may have done this less willingly or not at all (Pennycook, 1989). EFL teachers, especially if they themselves share the students' home culture, have had less need to examine what multiculturalism or diversity may mean for their classrooms; nonetheless, as language teachers we assume that most in our profession share at least some interest in and knowledge of other cultures and apply these understandings to their pedagogical practice.

That has not always been the case with U.S. educators from other subject areas. I return to the general area of education because this is where the original calls for alternative assessment in the U.S. originated, not in the fields of TESOL. I call

attention to this fact again to forestall an ahistorical analysis of alternative assessment and its role in the acts of teaching and learning. Though it certainly is useful to define terms and lay out theoretical underpinnings of alternative assessment, I would first like to broaden the notion of its relationship to the socio-cultural contexts which both precede it--and therefore give birth to it--and surround it, even now.

Early Beginnings: Teacher Education; Minority Learners

Early work in alternative assessment, when loosely defined as any form of assessment which does not take the form of paper and pencil, multiple choice tests, focused on teacher education programs (Mayher, 1990; Medley, 1979; Popham, 1974). Dissatisfaction with the quality of teachers coming out of education programs led to demands that broader, competency-based standards be applied to all educational programs.

During the same period, the first calls emerged for alternative assessment as a partial answer to the challenges of educating minority children, including African-American children and other children who speak English as a first language but may come from different cultural backgrounds. Educators had begun to feel that psychometric testing (paper and pencil tests with multiple-choice answers, standardized and tested for reliability and validity) do a disservice to certain groups of students. As one educator describing a project on alternative assessment exhorted, "Urban minority children have been particularly ill-served by the traditional paradigm of psychometric testing. The need for better assessment techniques is urgent" (Schwartz, 1979, p.1).

MORE THAN FORMAT CHANGES

A major difference between the teacher-oriented reforms and those targeted at minority-language populations was the intended purpose. Changes in standards for teacher education were focused on providing greater accountability for teacher performance; those for minority children were focused on providing fairer means of assessing what, in a phrase that has become commonplace in assessment discussions, learners "know and can do" (O'Malley & Valdez Pierce, 1996; Rothman, 1995). These two divergent purposes for creating alternatives to standardized tests have come together in the drive for modifying common educational practice. As O'Malley and Valdez Pierce (1996) observe:

> During the past 15 years, two themes have dominated public discourse on education in North American: (1) the perceived need for public schools to demonstrate more accountability to the societies that fund them, and (2) ways in which schools should accommodate to the rapid growth in cultural and linguistic diversity in urban centers and, specifically, ways of addressing the persistent educational underachievment of many students from culturally diverse backgrounds. (p. iii)

These two intertwined purposes, however, do not necessarily always complement each other. If alternative assessment is defined as a movement away from a prescriptive, rote kind of testing that de-emphasizes complex skills (Calfee, 1994; Herman, Aschbacher, & Winters,1992; Wiggins, 1993; Wolf, Bixby, Glenn, & Gardner, 1991), then some of the competency based, accountability orientation of "alternative" assessments may actually be in opposition to the need to recognize the ways in which learners may actually interact with the learning experience.

Let me return to the teacher education model to illustrate this point. John Mayher, a teacher educator of 25 years experience, relates the struggles in his department in the

early 1970's to fulfill the state supported recommendations for Competency Based Teacher Education (CBTE).

> CBTE was also sometimes referred to as PBTE (Performance Based Teacher Education) because the goal was to specify those things that teachers had to be able to *do* in order to succeed in classrooms. This alternative label revealed more clearly the behaviorist/positivist roots of the enterprise, since the criteria for success were all to be specified in observable terms. This led some early developers of such programs to look carefully at what successful teachers did and then write down lists of all of the observable behaviors they manifested. . . . What had been lacking . . . in the conceptual collapsing of Competency and Performance Based standards for judgment was an understanding of the distinction between competence and performance in an almost exact parallel to that developed by Chomsky for the study of language. That is, competence is a mental state which underlies and is enacted in performance, but because the structure of competence is so complex, as are the contexts which influence and affect how it is employed, that *performance itself does not unambiguously reflect or reveal competence*. One cannot, therefore, determine competence by *only* observing performance. (pp. 278-279, emphasis in original)

I have taken the liberty to quote at length here because I believe that Mayher's reflections on some of the manifestations of "alternative" assessment illustrate a critical point; the theoretical constructs behind any set of assumptions about assessment are just as important as the actual labels used. To put it in another way, the term "alternative assessment" in all its ambiguity, has been variously defined in one of the two rationales behind changing assessment practice: providing accountability and accommodating diverse learner needs. The first of these strands, in its most positive sense, means making standards clear so that learners know what is expected of them and can identify the traits which exemplify a strong performance; in the least learning-oriented sense, it means defining a complex learning experience into discreet, limited sets of behavior which are assumed to be the only demonstrations of understanding a complex act like teaching. In other words, just because the tests are no longer multiple choice does not mean that they do not rely on a view of education that sees the learner as a passive recipient of knowledge who demonstrates her/his acquisition of this knowledge by successfully completing certain pre-set acts.

THEORETICAL CONSTRUCTS

In a more central consideration of alternative assessment, it is not merely the absence of prescribed multiple choice tests, for example, that signals an alternative approach to assessment. I would argue that fundamental questions of the nature of knowledge and reality are at the heart of this discussion. Without considering what it means to *know* something, we can not posit that we have evidence of learning.

Feminist scholarship speaks to the question of what it means to *know* things. Maxine Greene (1993), a prominent educational scholar, cautions us to remember that models of knowing, of education, and, inadvertently, of assessment are not apolitical systems.

> Thinking of curriculum, realizing that it always emerges out of an interplay among conceptions of knowledge, conceptions of the human being, and conceptions of the social order, I want to lay stress once more on the way in which universals are structured . . ., categories are invented, and discourse is manipulated. (pp. 215-216)

The categories or the words we use to frame the systems we describe are made by human beings *for* the human beings who made them. It is all to easy to accept dominant modes of discourse as the "one, true reality," rather than as one, posited theory of reality.

Belenky, Glinchy, Goldberger, & Tarule's (1986) often cited research about ways women make sense of the world portrays two diverging views of knowledge: "received" knowledge (that information which comes from outside of the self) and "subjective" knowledge (the inner voice that a person hears which does not depend on authority figures for validation). In other words, knowers of received knowledge believe problems have one right answer. "Subjective" knowers believe that problems are contextual and may have multiple solutions depending on the context.

Belenky, et al.'s research illustrates the need to recognize more than one way of seeing, to acknowledge that a dominant model of education can do damage to those who are not privy to or do not agree with the accepted norms. More succinctly, education should confirm more than one type of human experience; yet, it has done little to look at, much less to accommodate, the ways of knowing girls, women, and other groups may have. As Balliro (1993) asserts, alternative assessment is not just about a different way of doing things, it is also about a different way of seeing things.

> A variety of "alternatives" are associated with the alternative framework: ethnographic as an alternative to the psychometric tradition; literacy-as-social-practices as an alternative to literacy-as-discrete-subskills; the acceptance of many "literacies" as an alternative to one privileged, monolithic view of literacy; and a preference for primarily qualitative data as an alternative to reliance on quantitative data alone. (p. 559)

Alternative assessment seeks to acknowledge a broader range of human interpretations of reality.

KNOWLEDGE: OBJECTIVE OR SUBJECTIVE?

Of course, the contention that reality might not be singular and universally recognized may be uncomfortable for many people, language teachers included. In fact, those who strongly believe in certain "objective" truths may turn to the sciences, math, or other supposedly objective fields of inquiry to validate their beliefs in one shared system of understandings. In his now classic text on scientific revolutions, Kuhn (1962), however, seeks to disabuse this notion of a singular reality by charting the ways in which science is also based on philosophical perceptions. Crudely distinguished, two polar views of reality can be seen in scientific inquiry: those based on an idea of objective reality which is knowable; the other based on a subjective view of the world which is context dependent. These differing world views, Kuhn proposes, in turn lead to different ideas about which problems to investigate and how to go about investigating them.

These differing views of knowledge can be traced to other sciences as well. Werner Heisenberg (1965), in analyzing research in physics, suggests that existing scientific concepts reflecting a positivistic tradition can only get at a small segment of reality. He proposes instead that understanding must come through natural language because the current traditions are too "rigid" a frame for the larger concepts of our language such as "mind, of the human soul or of life" (p. 207).

These two oppositional views might be more clearly explicated with a metaphor. The positivistic, quantitative tradition has often been associated with a "transmission" model of education (what Gardner, 1983, p. 6, calls a "uniform view" of schools), in which knowledge is perceived as a thing which exists in pure form external to people. The teacher/scholars who study the thing and best remember its qualities are, then

capable of most efficiently passing the thing on to novice learners (students). In this model the best teachers are the ones who know the most, who can pass on the most information. Learners are then passive objects, something like empty cups, waiting to be filled. The best students are those who can take in the most information *in the form in which it was transmitted.* This model disavows learning that occurs outside of the classroom or the presence of teachers because, as Mayher (1990) states, "positivists believe that "If something has been learned it must have been taught and if something is to be learned it must be directly taught" (p. 111). Put in another way, this view of learning suggests that "Individuals know nothing until--or unless--they take up the particular brand of formal knowing prescribed by the school" (Wolf, Bixby, Glenn, & Gardner, 1991).

Grant Wiggins, one of the central figures in discussions of alternative assessment, suggests that it is false to posit that if a person knows something, she or he knows it in any context. Instead he asserts that knowing is context dependent and, consequently, that learning has more to do with "good judgment of a disposition than the possession of information" (p. 202). More succinctly, it might be helpful to think of knowing in this alternative mode as "understanding" rather than just "memorizing" or "repeating."

Knowing and Cognitive Learning Theory

Cognitive learning theories provide us with more precise definitions of what it means to know and learn. As outlined in Herman, Aschbacher, and Winters (1992), the "discrete behavioral objectives" that required the mastery of "rote basic skills" prior to "complex thinking skills" have been contradicted by findings from contemporary cognitive psychology. Cognitive learning theories provide an alternative vision of education in which learning is an *active* process which is not linear, learners have diverse talents and ways of approaching the learning process, and learners benefit from reflecting on their own learning, as well as working with others (pp. 14-17).

Educator and cognitive psychologist Robert Calfee (1994) traces the more traditional, positivistic transmission model of education in the U.S. to a link between Educational Administration and Behavioral Psychology in the early 1900's. In this model, schooling was viewed from an efficiency perspective and learning was thought to take place in a step by step mode, something like a production line. In fact, Calfee terms the schools in this educational model "factory schools." Although Calfee observes that an alternative view of learning was being explored as early as 1963, the work of these early cognitive psychologists had no immediate impact on testing. Their work was largely kept within their own discipline of psychology and behavioristically oriented, objective tests were already very much a part of schools. It has only been recently that work in this area has been explicitly connected with the learning that takes place in schools.

Calfee summarizes three of the major assumptions of the cognitive model as they apply to curriculum.

- The mind is a living organ that depends on purpose and coherence, not a warehouse to be filled with information.
- Reflective learning built on genuine dialogue and social interactions is more long-lasting and transferable than rote acquisition.
- Previous experience is essential for effective learning (p. 8)

Cognitive learning theory calls into question the notion that something is known or has been learned just because it has been *remembered.* Or, as Wolf, Bixby, Glenn, and Gardner (1991) observe, "These alternatives might permit the assessment of thinking

rather than the possession of information. . . . However . . . this transition will have to be wholesale because . . . the conduct of standardized testing and its alternatives are in no way superficial matters or mere surface features. They derive from radical differences in underlying conceptions of mind and of the evaluation process itself" (p. 33). It is not enough to simply add alternative assessments on to a traditional classroom; a true examination and use of alternative assessment requires a re-examination of curriculum and theories of knowledge--a new way of seeing.

CRITICISMS OF ALTERNATIVE ASSESSMENT MODELS

This need for revision--literally a new way of seeing--may help explain why alternative assessments, even with the backing of substantial research into the human learning processes and the accompanying pedagogical practices to support such a learning model (such as Communicative approaches, Whole Language, as well as Cooperative Learning), alternative assessment has been the recipient of many attacks. While I do not wish to downplay the need for a rigorous, thoughtful examination of the assessments we develop, I would like to address an issue that I think is central to the comparative vulnerability of alternative assessment to attacks from those who support more traditional testing models.

Huerta-Macias (1995) notes that three issues seem to be at the center of questions about alternative assessment: validity (whether a test/assessment actually tests the thing which it claims to), reliability (whether the results of a test/assessment would be the same over a period of time, given all other things being equal), and objectivity (whether the test/assessment is without biases/prejudices). She summarizes: "Proponents of alternative assessment do not suggest that we overlook these criteria, for any high quality assessment must adhere to them. Rather, the suggestion is that we apply new words that have been borrowed from the literature on qualitative research" (p. 9). Huerta-Macias goes on to talk about "trustworthiness" (as used by Guba & Lincoln, 1981) and "triangulation" (Patton, 1987). Her discussion of alternative labels for good assessment criteria provides a jumping off point for more explicit discussions of what standards should apply in the making of alternative assessments.

Redefining Standards: Avoiding the Danger of Co-optation

I have argued previously (Butler, 1995) that a danger exists in so easily defining the value of any research (read here as assessment or learning) mode by the standards of another, more universally accepted one. As with the earlier discussion of different ways of knowing, losing sight of the distinctions does a disservice to both. It is the wrangling interaction that calls us to forego laxity. Nonetheless, people thinking about alternative modes should also be aware of the co-optive possibilities of too often, too carefully judging one mode by the criteria of another. As Smith and Heshusius (1986) argue, a shift toward synthesis may actually signal the subsuming of qualitative (alternative) forms by the dominant paradigm (quantitative, traditional tests).

> During the last few years the quantitative-qualitative issue has gone from a situation of conflict to one of compatibility and cooperation. The contention of this paper is that this transformation is not based on the development of a legitimate *via media* between the two perspectives, but is actually a matter of the "capture" of qualitative inquiry by the quantitative approach. (p. 10)

They see evidence of this capture in the use of quantitative criteria to evaluate qualitative research.

Of course, other qualitative researchers (Guba and Lincoln, 1989; Mishler, 1990) have made this same point. Mishler proposes that "Criteria and procedures based on the dominant experimental/quantitative prototype are irrelevant to these studies in the literal sense that there is nothing to which to apply them. When the standard model is misapplied, as it often is, inquiry-guided studies fail the test and are denied scientific legitimacy" (p. 435). Papagiannis, Klees, and Bickel (1982) make a similar point but in relation to the adoption of educational innovations, a particularly apt analogy for this discussion. They propose that educational innovations may fail to take hold because their suitability has been gauged in terms of the dominant majority, failing to take into account the perspectives of those who will use the innovations. Kamarck Minnick (1990), writing from a feminist perspective, proposes that claims to knowledge suffer from this same kind of political frame in which the majority definitions reproduce themselves and become the standard against which all else is judged. She suggests that, "It is circular reasoning when what has been accepted as knowledge is considered as definitional of or essential to knowledge itself, a move that disguises partiality as impartiality" (p. 152). She further observes that all claims to knowledge are necessarily constructions rather than "found" truths; therefore, no one knowledge system can have innate authority over another. This political interpretation of power relations between the two approaches is also echoed in Reinharz (1990).

An alternative to any co-opting or subsuming of one framework by another, however, would be a mutually beneficial dialogue which would force proponents of both research (assessment) paradigms to examine their own assumptions more closely, leading to a greater refinement of research methods, techniques, and beliefs. Such refinements would contribute to the quality of the research (assessment) conducted rather than preventing meaningful research (assessment) from being done. As Allwright (1983) puts it, such a dialogue ". . . performs the useful service of helping keep all of us a little more on our toes about what we are doing and what we can sensibly claim about our results" (p. 201)

Put in another way, as long as we use a dominant mode, in this case traditional psychometric testing theory, to inform our explorations of alternative assessment rather than to define them, we will more fully ensure the health of both.

THE NEED FOR RIGOR IN DEVELOPING NEW ASSESSMENT MODELS

This admonition to stay true to the theoretical and philosophical underpinnings of alternative assessment is, however, not a call for looseness or lack of reflective insight when designing alternative assessments. It is the reverse. I believe that it takes even more rigor to define the parameters of an assessment when it is not simply being made a shadow of more traditional tests. Just as it is easy to put together a bad multiple choice paper and pencil test, it is easy to put together a bad alternative assessment. The reverse is true as well, and the task at hand in making alternative assessments is made all the more difficult by the lack of rules, the still relative scarcity--even with the preponderance of publications--of good models that apply to the diverse situations for which teachers may wish to use these assessments. Several good guidebooks do exist (Herman, Aschbacher, & Winters, 1992; Marzano, Pickering, & McTighe (1993); O'Malley & Valdez Pierce, 1996), and Angelo and Cross (1993) provide a wealth of ideas for ungraded, classroom based assessments, yet, because of the very nature of alternative assessment, it is hard to apply assessments "as is" to classrooms. Assessments are context dependent; subsequently judgments of what assessments might be appropriate require substantial thought on the part of individual teachers.

Armed with some understanding of the historical, social, political, and theoretical underpinnings of alternative assessment, classroom teachers as well as researchers and administrators should be all the more prepared to face the task ahead of them. As

Ewell (1991) notes, "At its extreme, the impetus toward more naturalistic designs can be seen as sanctioning a methodological philosophy of 'anything goes'" (p. 110). If educators are to be successful in forging new understandings of what it means to understand and evaluate student progress, then they will need to respond to the challenge to understand and think through a set of criteria on which the relative usefulness of alternative assessments may be judged. As Wolf, Bixby, Glenn, and Gardner (1991) observe, "Whereas there is considerable criticism of the approaches taken by standardized tests, as yet we have no such critical tradition for new modes of assessment. And we cannot be without one" (p. 60).

REFERENCES

Alderson, J. C., Clapham, C. (1995). Assessing student performance in the ESL classroom. *TESOL Quarterly, 29* (1), 184-187.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Allwright, D. (1983). Classroom-centered research on language teaching and learning: A brief overview. *TESOL Quarterly, 17*, 191-204.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Balliro, L. (1993). What kind of alternative? Examining alternative assessment. *TESOL Quarterly, 27* (3), 558-561.

Belenky, M. F., Glinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind.* New York: Basic Books.

Brindley, G. (Ed.). (1994). *Language assessment in action.* Sydney, Australia: NCELTR/Macquarie University.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Brown, J. D., & Yamashita, S. O. (Eds.). (1995). *Language testing in Japan.* Tokyo: JALT.

Butler, P. (1995). Looking in and seeing out: A qualitative study of an adult ESL classroom. Unpublished doctoral dissertation, New York University, New York.

Calfee, R. (1994). *Implications of cognitive psychology for authentic assessment and instruction* (Tech. Rep. No. 52). Berkeley: University of California, National Center for the Study of Writing.

Census results show dramatic changes. (1990, March 11) *New York Times*, p. A, 1.

Dewey, J. (1934). The need for a philosophy of education. *The New Era in Home and School, XV*, 211-214.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Higher Education, 17*, 75-125.

Garcia, G. E., & Pearson, P. D. (1991). The role of assessment in a diverse society. INE. F. Hiebert (Ed.), *Literacy for a diverse society* (pp. 253-278). New York: Teachers College Press.

Faltis, C. (Ed.). (1995). Alternative assessment [Special issue]. *TESOL Journal, 5* (1).

Faltis, C. (Ed.). (1996a). Self-assessment: Guiding learners to assess their strategies and abilities to communicate. *TESOL Journal, 5* (3).

Faltis, C. (Ed.). (1996b). Learning styles and strategies [Special issue]. *TESOL Journal, 6* (1).

Garcia, G. E., & Pearson, P. D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education* (pp. 337-391. Washington, DC: American Education Research Association.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Genesee, F. (1996). *Classroom based evaluation in second language education.* Cambridge: Cambridge University Press.

Green, M. (1993). *Diversity and inclusion: Toward a curriculum for human beings.* Teachers College Record, 95 (2), 215-221.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Guba, E. G. & Lincoln, Y. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Hancock, C. (Ed.). (1994). *Teaching, testing, and assessment: Making the connection.* Lincolnwood, IL: National Textbook Company.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Hicks, M. (1995). *Learning to know: The multi-cultural self and liberal education.* Unpublished manuscript. Teachers College, Columbia University, New York.

Hill, C., & Parry, K. (Eds.). (1994). *From testing to assessment: English as an international language.* London: Longman.

Homlar Fradd, S., Larrinanaga McGee, P., & Wilen, D. K. (1994). *Instructional assessment: An integrative approach to evaluating student performance.* Reading, MA: Addison-Wesley.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5* (1), 8-11.

Knefelkamp, L. (1992). The multicultural curriculum and communities of peace. *Liberal Education, 78* (2), 26-35.

Kuhn, T. (1963). *The structure of scientific revolutions.* Chicago: The University of Chicago Press.

Law, B , & Eckes, M. (1995). *Assessment and ESL.* Winnipeg, Canada: Peguis.

Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model.* Alexandria, VA: Association for Supervision and Curriculum Development.

Mayher, J. S. (1990). *Uncommon sense: Theoretical practice in language education.* Portsmouth, NH: Boynton Cook.

Medley, D. M. (1979). Alternative assessment strategies. *Journal of Teacher Education, 29* (2), 38-42.

McGrane, B. (1989). *Beyond anthropology: Society and the other.* New York: Columbia University Press.

Minnich, E. K. (1990). *Transforming knowledge.* Philadelphia, PA: Temple University Press.

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*, 60, 415-442.

O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers.* Reading, MA: Addison-Wesley.

Papagiannis, G. J., Klees, S. J., & Bickel, R. N. (1982). Toward a political economy of educational innovation. *Review of Educational Research*, 52, 245-290.

Patton, M. Q. (1987). *Creative evaluation.* Newbury Park, CA: Sage.

Pennycook, A. (1989). The concept of method, interested knowledge, and the politics of language teaching. *TESOL Quarterly, 23 (4)*, 589-618.

Philipson, R. (1992). *Linguistic imperialism.* Oxford: Oxford University Press.

Popham, W. J. (1974). Minimal competencies for objectives-oriented teacher education programs. *Journal of Teacher Education, 25* (1), 69-73.

Reid, J. (Ed.). (1995). *Learning styles in ESL/EFL classrooms.* Boston: Heinle & Heinle.

Reinharz, S. (1990). So-called training in the so-called alternative paradigm. In E. Guba (Ed.), *The paradigm dialog* (pp. 290-302). Newbury Park, CA: Sage.

Rothman, R. (1995). Telling the story of: How you communicate with stakeholders can make or break reform efforts. *School Administrator, 52* (11), 6-11.

Rudner, L. M., & Boston, C. (1994). Performance assessment. *The ERIC Review, 3* (1), 2-12.

Spolsky, B. (1995). *Measured words.* Oxford: Oxford University Press.

Schwartz, J., L. (1979). *What do we know about teaching and learning in urban schools? Volume 3: Assessment that respects complexity in individuals and programs.* (Report No. UD020353). Washington, D. C.: National Institute of Education. (ERIC Document Reproduction Service No. ED 185 159)

Smith, J. K. & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher,* 15, 4-12.

Spolsky, B. (1995). Measured words. Oxford: Oxford University Press.

van Troyer, G., Cornwell, S., & Morikawa, H. (Eds.). (1995). *On JALT 95: Curriculum & Evaluation; Proceedings of the JALT 1995 International Conference on Langauge Teaching/Learning.* Tokyo: JALT.

Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan,* p. 200-214.

Wolf, D., Bixby, J., Glenn, J. III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Higher Education, 17,* 31-74.

Part 2

Classroom Applications

Chapter 3

Portfolio-Based Assessment

MARICEL SANTOS

PART ONE

Teaching, learning, and assessment: Self-discovery through portfolios

ßßß

We meet awkwardly.
I invite you to walk.
I find you dancing.[1]

ßßß

INTRODUCTION

My work in portfolio assessment has had a profound effect on me and my teaching; it has made me a firm believer in the idea that learners have a right to ask, *why do I learn what I learn?* I believe that while we don't always know how to respond to this question, the search for answers allows learners to reclaim their own learning processes. As a graduate student, I created a portfolio which documents some of the answers I found. The portfolio reveals what I feel and believe about myself and about learning. I am convinced that my learning is more valuable and meaningful because I have exercised the right to ask this question. Now, I hope that by helping students in my English classes to complete their own portfolios they experience the same sense of self-empowerment.

Assessment is inextricably bound to what those intimately involved in the process *value* about learning and their roles in the classroom. Portfolios encourage students and teachers to make these values explicit so that all classroom participants are more aware of their own and one another's expectations about learning and classroom roles. Sweet (1993) writes that "research shows that students at all levels see assessment as something that is done to them on their classwork by someone else. Beyond 'percent correct,' assigned letter grades, and grammatical or arithmetic errors, many students have little knowledge of what is involved in evaluating their classwork." Portfolio assessment represents an opportunity for both students *and teachers* to inform themselves about assessment; teachers are more in control of assessment practice and students take more responsibility for their learning. Portfolio decisions and guidelines come from the students and teacher, and as a result, assessment information is more meaningful to those who are most directly influenced by its results.

This chapter pieces together journal excerpts, student portfolio writing, and introspective thoughts about my experiences with portfolios in a class of twenty-eight

[1] Anonymous reflective piece entitled "Teaching" in Women and Teaching by Maria Harris (1988, 4).

university English learners. In the first part of this chapter, I will describe the impact of portfolio assessment on our classroom in two areas :

- how the portfolio enhances learner responsibility and motivation, and
- how the portfolio provides teachers with rewarding and constructive feedback about the effectiveness of their teaching and about student performance.

In the second part of the chapter, I will provide more detailed, hands-on description of our portfolio framework.

The rewards of portfolio assessment result from the way the entire class approaches the portfolio task: the teacher and the students are focused on learning and the search for student work which demonstrates learning progress; students are responsible for creating a portfolio and making judgments about their own work; the teacher must work to create a learning environment where students are encouraged to reflect and discuss their opinions about learning; the teacher is also responsible for reviewing and commenting on student work, and meeting with students to guide them in their portfolio creation.

My experiences confirm that portfolio-based assessment is rigorous and demanding, yet also uplifting and triumphant. Working with the students to complete a portfolio was an awkward meeting of minds, both of us a little unsure of one another's actions and intentions. The students for the most part were accustomed to standardized testing formats and did not readily understand the portfolio project. I had a vision of the general portfolio process because I had completed one myself. However, as I tried to guide the students through the analysis and organization of their work, I often struggled to find the words to explain the portfolio purpose and procedure. There were many questions, such as: *What should be put in the portfolio? How will the portfolio be evaluated? Are there requirements everyone should fulfill?* These difficult questions were answered over time and often collaboratively.

The quality of the students' portfolios surpassed my expectations. Portfolios made assessment a more engaging process, and the assessment information taught me a lot about learning and my role as a teacher. By asking the students to complete a portfolio, I had invited them to learn more about themselves and learning, and in the end, the portfolio led us all down the path of self-discovery.

FOSTERING LEARNER OWNERSHIP AND RESPONSIBILITY

In its simplest form, the portfolio is a collection of student work. One of my students noted that a portfolio can sometimes be merely "a folder of all my papers." Initially, many students viewed the portfolio project this way. When students recognized, however, that the portfolio represented a genuine opportunity to display their individual areas of language progress, interests, and opinions, the portfolio became more personalized and therefore more informative about their performance and attitudes. Tierney, Carter and Desai (1996) assert that portfolio-based assessment purposely aims to highlight learner individuality, so that the relationship between a learner's goals and classroom instruction is more clearly observable. Evidence of this individuality can be seen in the various ways the students viewed their portfolio product. Here are some student definitions of a portfolio:

> *A collection of what I did in this semester. I will be able to look back easily (on) what I did in this class in the future.*

> *I think a portfolio is to review what I studied. We are tend to forget our mistakes we made. So, we study our mistakes again by review a portfolio.*

A balance sheet or check sheet of a student. (Its purpose) is to affirm why we are studying English.

To make a student realize at what point he/she has come to be a successful learner and what he/she must do next.

Gottlieb (1995) emphasizes the personalization of the portfolio when she says that these "collections are an expression of the students, their lives, and their identities. Ownership resides in the students, and with the guidance of the teacher, they should have flexibility in shaping the portfolio and access to the contents" (p. 13).

In the beginning of the semester, I provided the student with guidelines for the portfolio content, and the criteria by which their portfolios would be judged. I asked them to store all their classwork in one place, in a folder or box. Periodically, over the semester, we talked as a class and in student-teacher conferences about these guidelines and criteria, clarifying them so that we better understood how portfolio quality would be judged. Also, throughout the semester, I asked the students to sift through their growing collection of classwork and reflect on the significance of each item. They asked themselves questions like: *What did I do to complete this work? What did I learn? How does this work demonstrate my English skills are improving or changing? Is this work an achievement for me? Am I happy with its results? How might I revise or improve this work?* Towards the end of the semester, this sorting process became more critical as the students began narrowing down their few choice selections to put in their portfolio.

Calfee (1994) writes that some chosen works "are assigned, others are self-initiated. Some are substantial projects, others a page or less. Each individual assembles his or her own folder, but many of the projects are collaborative" (p. 3). Sometimes, students form small groups and brainstorm about possible items to include in the portfolio: TOEFL score reports, a video of a student's speech, or a computer print-out demonstrating the student's ability to use the internet. Many works demonstrate stellar language achievement. In exemplary situations, a student will revise a less-than-stellar writing assignment and include the first and second drafts as evidence of improvement over time. One of my students included a quiz she had failed to illustrate that language learning involves making mistakes and learning from them.

Students are responsible for selecting work and also for writing *cover notes*[2] -- the written analyses which accompany each work and explain why the student chose to include the item in his or her portfolio. Writing these brief reflective pieces promotes valuable habits of self-analysis in the students, because the task asks them to do the following:

- to identify areas where genuine language learning has occurred,
- to explain weak performances and indicate areas where they need more help, and
- to describe real English learning needs and viable learning goals.

As a result of this reflective work, the students realize that the portfolio represents more than just *a folder of all the student's papers.* The portfolio serves as an *integrating vehicle* in students' language learning. Students are asked to make connections

---

[2] *Cover note* and *voice* are terms used by the faculty and students at the Monterey Institute of International Studies who use portfolios to evaluate teachers in training. The guidelines and criteria I provided the Japanese students are adaptations of what I learned about portfolios while studying at the Monterey Institute. I would like to thank Professor Peter A. Shaw of the Monterey Institute for his advice and support during the evaluation stages of the portfolio process.

between their efforts and consequences of their actions; they make links between learning goals and classwork which demonstrates they are working towards these goals.

As I read through the twenty-eight portfolios, I was impressed with the amount of energy and creativity the students had devoted to the portfolio project. The high quality of the students' portfolios confirms Calfee's opinion (1994): "The bottom line is that, if portfolios are taken seriously, most students react seriously" (p. 6). When I first introduced the portfolio to the students, I told them that one, a portfolio was required to successfully complete the course, and two, the process would require them to take seriously the responsibility of creating the portfolio and reflecting on their choices. Were the students able to handle so much responsibility for their learning and assessment? Portfolio-based assessment assumes that students *can and should* accept this kind of responsibility. Calfee (1994) takes this perspective when he labels the learner the "assessment expert" in the portfolio framework (p. 7). In a similar way, Paulson (1993) asserts:

> You can't separate learning from assessment. They occur simultaneously. In fact, I believe that the most effective instruction occurs when the learner and assessor become the same person. If we want students to take charge of their learning, they must also take charge of their assessment. (p. 2)

Nurturing self-directed learners takes a long time, no doubt much longer than the duration of our portfolio process. I believe that for many of the students, the portfolio served as a catalyst to this realization of learning ownership. The words of one student, Sayaka[1], illustrates the point I am making here. Sayaka writes in the self-evaluation section of her portfolio:

> *I think I was being unwilling to (try) new things. When I first heard about the portfolio, I was not too excited about it. Honestly, I did not want to do it at all. But as I read other classmates' portfolios, I felt like I was told the real purpose of the portfolio... to be the reminder of my goals and dreams. As I read my portfolio, I realized even though I got good grades for my work, I did not have any purpose to them. I was wasting my efforts. Now (that) I realize I was mistaken, I would try hard to improve my skills and have a real "achievement" next semester...Maybe I achieved something by realizing that I wasted precious time (this semester).*

I believe that the portfolio experience represents an important stage in Sayaka's process of learning, one in which she is discovering more about what she is capable of achieving in English.

Others students wrote about feeling guilty for not trying harder when the results of their efforts were not what they had expected. Reflective tasks for the portfolio encouraged them to identify obstacles in their learning and quickly begin considering solutions, rather than blaming themselves, the teacher, or course activities for their apparent lack of progress. Here is an essay excerpt written by Keiko who describes how she was able to take a more positive view of her English speaking ability:

> *As for me, I achieved two big goals in (this class). First, I was not ashamed to speak in public anymore. Before doing our project, Quiz Show,... I was so shy that I felt embarrassed about making a remark in public.... But I didn't like such a personality and wanted to change it. After the presentation, I felt I did it! Since then, I became confident in myself... This is the example of achievement coming from satisfaction and fullness. The other is an example of getting over my weak points. I was poor at speaking and listening skills before... But I could improve these skills little by little through many group discussions, reading discussions, and*

[1] The names of students have been changed.

other activities. Especially, project work gave me a chance to say my opinion without any hesitation. I'm really glad that I can say what I want to say because it was one of my goals...

Keiko put this essay in her portfolio, in addition to work from an in-class oral presentation and materials from group discussions when she was a leader. I was impressed with Keiko's portfolio because she was able to choose portfolio work which nicely complemented her reflective thoughts.

I was also impressed with the ideas of another student Hiro who wrote about two kinds of achievement. Hiro wrote that one kind of achievement is "*based on others people's evaluations... It refers to the kind of achievement whose main purpose is getting good evaluations from other people.*" The other kind is "*based upon one's satisfaction... it means the achievement which respects other people's evaluations but also one's own thoughts.*" Hiro wrote that the latter kind is more useful because it takes into account what both teachers and learners believe about the standards for quality work. For his portfolio, Hiro set a personal standard for achievement: the work must display his individuality. He included a video commercial because it best illustrated his creativity, and talked about how his listening skills had improved as a result of individual effort. Like Keiko, Hiro demonstrated that the students' own personal criteria could be used to evaluate their classwork.

Hiro also wrote that "*true achievement (means being) satisfied to the full without being puzzled by others' opinions.*" Hiro's portfolio reminds us that the process and products of assessment should not be confusing, an insight echoed by Yuri in her portfolio:

This portfolio is all my brain.... In this portfolio, I understood the things very well that we studied in (this class). And I understood which part I understand and which part I don't understand.... I realized that grade is not a point but an effort. And at least I think that looking the portfolio makes us know how much effort each person does to(create) a portfolio.

The value of Yuri's portfolio is clear. It enabled her to more accurately assess her own English progress and understand instructional aims. Moreover, her words send out an insightful message about the relationship between learning and assessment: language learning results from personal effort, and assessment is more valuable to the student when it deliberates language performance as an individual achievement.

FOSTERING PROFESSIONAL AND PERSONAL GROWTH

One of the most lasting impressions of reading students' portfolios is that I felt privileged to have access to their introspections and insights about their work. The information in their portfolios provided me with extensive feedback about the effectiveness of my teaching. Reading their reflections prompted me to reflect on ways I could improve course activities or further aid independent learning. I learned more about the various proficiency levels and learning styles in my classroom, and as a result, I could better group students for collaborative learning tasks. I was able to see where students needed more skillwork and adjust course activities to match these needs.

Hughes (1989) makes a good point when he says "the proper relationship between teaching and testing is surely that of partnership...(testing) should be supportive of good teaching and, where necessary, exert a corrective influence on bad teaching. If testing had a beneficial backwash on teaching, it would have a much better reputation amongst teachers" (p. 2). I was more interested in the assessment process because I was confident that the process and the results would be useful to my classroom teaching. I think that the partnership formed through portfolios, however,

offers more benefits than Hughes mentions here. The portfolio system can boost teaching enthusiasm since it rewards the teacher with positive feedback when successful teaching takes place. The portfolio process enables teachers to free themselves from conventional boundaries of one-way (teacher-to-student) instruction, away from pressures to rank or give grades. Teachers also gain an opportunity to rethink their assumptions about learning and assessment. With portfolios, assessment promises to be the "type of assessment that is integral to instruction" (Valdez Pierce 1996, p. 9), and no longer the end-of-the-semester burden. Teachers are able to avoid settling for quick-fix solutions and move beyond "one-right answer" perspectives to assessment. In short, one area of powerful influence of portfolio-based assessment on teaching is this: the students are encouraged to view themselves as responsible learners, and therefore they work responsibly. As a result, the teacher feels more motivated to teach.

Research by Sheingold, Heller and Paulukonis (1995) reveals that "when teachers, in collaboration with their colleagues, spend time thinking about, looking for, and evaluating evidence of student learning in the context of activities that take place in their classroom, there is a significant opportunity for them to grow" (p. 20). Portfolio-based assessment then represents a viable opportunity for teachers' professional and personal development, primarily because it engages teachers in critical dialogue about their role in the classroom.

Brookfield (1995), a strong advocate of critical reflective thinking in adult education, effectively summarizes the potential value of portfolio-based assessment to teachers and teaching:

> Knowing something of how students experience learning helps us build convincing connections between what we want them to do and their own concerns and expectations. We can make a stronger case for students to take seriously what we say. Researching students' perceptions of our actions and words alerts us to problems that our behavior is causing and to mistakes that we might otherwise miss. This, in turn, means that we can make more appropriate decisions about how and what to teach. (p. 93)

Applying Brookfield's thoughts to portfolio-based assessment, we can see that the portfolio is the vehicle for on-going classroom investigation and self-inquiry, ultimately placing the teacher and students in positions of "assessment expertise."

Through portfolio-based assessment, teachers are in a position to continue learning about themselves. This exploration arises from a need in teachers to affirm why they do what they do. Affirmation is a self-initiated, very personal process, and the portfolio is the framework which enables us to pursue this exploration, while also expanding our understanding of what our students are doing.

CONCLUSION

I am confident of the rewards of the portfolio as a process and product. The portfolio was one way of affirming that all of us, students and teachers, are best measured by our development over time. At the same time, I think the benefits of portfolio assessment were not fully realized in my classroom. Perhaps, the most disappointing part of the whole portfolio process was that my students and I were the only ones to read them in the end. In the future, more classroom research and institutional support would be needed to fully explore the potential of portfolios as an alternative assessment tool. More faculty input on the guidelines would undoubtedly improve my ability to present portfolio assessment to the students, to refine the evaluative criteria, and to better utilize the assessment information in the classroom.

While writing this chapter, I wondered, is portfolio-based assessment an appropriate tool for exploration at the Japanese university? For me and my classroom

of twenty-eight eager English learners, the answer is *absolutely*. Yet, one or two classroom successes alone will not sustain this exploration of new assessments. There are several possible problems concerning portfolio use in English classes at the Japanese university. Typically, Japanese English classes are large (40-50 students) and meet on a limited basis (sometimes only once a week), so the portfolio project might seem to be a logistical nightmare. Another possible problem is that Japanese students may not be prepared to handle the serious reflection tasks involved in portfolio creation. Davidson (1995) observes that the heavy emphasis placed on memorization for multiple choice exams partially denies Japanese students the chance to analyze or think deeply about serious topics in the classroom. It is very likely then that students would need extra time and much guidance while completing their portfolios, but still yet another problem is that Japanese teachers may not have the necessary time, outside their regular teaching schedule and administrative duties, to research a portfolio framework, prepare portfolio tasks, and meet with individual students.

Another more fundamental obstacle which may slow the exploration of portfolio assessment concerns the widespread, heavy reliance on standardized testing at many Japanese universities. Calfee (1994) observes that "standardized test are not likely to go away, but it does seem feasible and desirable to restore balance to the assessment of student achievement" (p. 9). Calfee's words suggest that the exploration of new assessments, like portfolios, is one way of bringing balance to assessment practice in language classes at the Japanese university. Instructors may gather necessary information (scores, rankings, statistics) from standardized test results, but portfolios, unlike any standardized test, will "document a story for every student" (Huertas-Macías, 1995, p. 10), thereby providing a more comprehensive -- and perhaps more accurate --picture of each individual learner.

The nature of portfolio-based frameworks are that they are highly contextualized, structured to fit the needs and purposes of each classroom. Clearly, there are no easy solutions to the problems the Japanese classroom may face, but the portfolio represents an opportunity for teachers and students in the Japanese university context to tackle these problems and gain a new and insightful perspective on assessment.

Paulson (1993) says that before embracing or discounting the portfolio, we must first ask ourselves "why we assess at all" (p. 2). I agree. This more fundamental question must *first* be explored by teachers and students in the classroom, and then their perspectives must be balanced with administrative and institutional concerns. This pursuit of answers is a difficult but necessary process, if our assessment methods are to provide useful and meaningful information about what goes on in our classrooms.

Part Two

Building a Portfolio Framework: Guidelines and Suggestions

First you see the hills in the painting;
then you see the painting in the hills.
Li-liWeng, Chinese painter[3]

My experience in using portfolio-based assessment shows that the process is highly contextualized; the process is focused upon individual performance and actual student work, so that learning goals, evidence of progress and achievement, and areas of instructional impact are made explicit and reflected upon by classroom participants. Just as no two portfolios are alike, portfolio approaches will also differ according to the curriculum needs and assessment purposes of each classroom. This part of the chapter describes the portfolio process which evolved over four months in my university EFL class with twenty-eight Japanese sophomores. I will explain the organization and contents of my students' portfolios, their reactions to the portfolio task, and some of the difficulties we encountered.

In an effort to cover a wide range of portfolio concerns in a limited number of pages, I have chosen to present the information in condensed forms: a question-and-answer section which touches upon issues of portfolio guidelines and management, tables and charts which outline key features of the portfolio process, and where possible, sample student handouts and classroom activities. This information aims to show the reader that the portfolio process involves seriously contemplated, systematic approaches to the documentation and analysis of student work.

The following pages provide only a glimpse of the kind of work and effort our class put into their portfolios. Ideally, readers would have the chance to examine some of the student portfolios and talk with the students and me personally about our experiences and reactions. Hopefully, these pages will help readers envision one approach to portfolio-based assessment and also prompt them to think of ways in which portfolios can be developed in other language learning situations.

As the opening quotation suggests, implementing new assessments (or any new method) in the classroom often begins with a few ideas and artifacts we gain from other classrooms. We build upon these details and by working with our students over time, we succeed in making the portfolio process our own.

[3] Poem quoted in Women and Teaching by Marta Harris (1988: 20).

Questions and (Possible) Answers

1. How do I get started?

One of the first steps is to determine your assessment purpose, in light of your own teaching goals and the curriculum agenda. Some possible purposes for assessment might be to determine advancement to higher levels, to evaluate students' abilities relative to curriculum standards, to exhibit students' accomplishments, or any combination of these. In my classroom, my primary assessment purpose was to learn more about my students' current language abilities and to encourage them to be self-reflective, independent learners.

Another first step is to determine the assessment information (the kinds of portfolio work) which will best fulfill your assessment purpose. You might consider what classroom behavior and work will best reflect instructional outcomes and desired student performance. For example, if you are using portfolios to evaluate oral skills, the portfolio might include a personal statement of speaking goals, speech notes and evaluations, checklists of conversational strategies, oral interview scores, or pronunciation tapes.

After determining purpose and the kinds of evidence you want to see, it is easier to visualize the portfolio process, specifically the series of activities which will guide the students through the organization and analysis of their work. This step raises many questions of portfolio management, such as:

How much classroom time should I set aside for reflection and portfolio discussion?

How much time can I spend meeting with students and responding to their regular classwork and their portfolios?

How can I integrate portfolio activities with the regular curriculum?

It may be useful to chart out your semester schedule, taking into consideration how much time you are willing and able to spend developing a portfolio framework. Talk with other colleagues to see if they are willing to collaborate on developing guidelines and evaluating the final products, so that the process does not place enormous burden (time and effort) on one teacher. If possible, keep a journal about your reactions to the portfolio process and frequently seek out feedback from colleagues.

2. What is the portfolio process?

The portfolio process can be viewed as a series of six major stages, as outlined below. (Also see handout B, *The Portfolio Process.*) These six stages have no finite beginnings and endings; rather, they flow together, often overlapping and merging. There is also no set time line for the duration of each stage or the entire process. Our portfolio process took 13 weeks to complete, most of which was spent on reflection and documenting the work which went into the portfolio. Students who do not have the language skills to communicate readily their reflections in English may need a lot more time and guidance to analyze their learning and work, thus I imagine the project could easily be extended over an entire year.

Six Major Stages in the Portfolio Process

PREPARATION	Teachers reflect on assessment purposes and classroom roles. With their students, they begin to visualize the portfolio process and to clarify criteria for judging portfolio quality.
GATHERING & SIFTING OF WORK	Students, with the guidance of their teachers, review their work and contemplate their final choices. Students share portfolio plans and ideas with other students. In groups, they discuss their possible choices and reasons for their decisions.
REFLECTION	Students, with the guidance of their teachers, work on articulating their reflections about their portfolio work in writing. They also spend time contemplating the value of the portfolio in their language learning.
CONSTRUCTION	This stage involves the physical piecing together of the portfolio. The students choose the notebook/box/folder where they will organize all their documents. They also decide how aesthetically appealing they wish to make the final product (e.g. using colored paper, adding multi-media items like tapes).
EVALUATION	Each portfolio is evaluated three times using the same criteria: by the portfolio creator, a peer and the instructor. The teacher also considers how the assessment information will be used to improve instructional activities.
SHARING	The entire class shares their portfolios and celebrates their completion of a challenging project. When possible, the teacher shares the assessment results with other teachers and administration.

3. What kind of student work goes into the portfolio? Who decides what work is chosen?

The portfolio contents will vary according to the assessment purpose as decided by you and your students. Early in the semester, I gave the students an outline of portfolio contents (see handout B, *Introduction to the Portfolio*). The outline provides the students with categories of work (e.g. Reflective Work, Academic Work), but it is primarily the students' responsibility to decide which work best fit these categories. Some works are required of all portfolios; for example, the Personal Essay on

Defining Achievement, and the three portfolio evaluations. Other areas (for example, the Personal Section) are completely open and varied according to the student's individual interests and creativity. To optimize the variety of possible selections, students, in groups and in individual conferences, discuss the kinds of work which could match the categories.

Based on the range of work included in my students' portfolios, I compiled this list of portfolio categories (in bold) and possible entries.

I. Personal Statements about own learning and beliefs about learning

A. Reflective Work: Student's beliefs about learning
Portfolio Introduction: overview of portfolio contents
Personal reflective essay (Topic: My Definition of Achievement)
Rubrics on characteristics of successful language learning

B. Self-Assessment Documents
Letter of self-introduction to the teacher
Needs analysis of language goals and interests
Statement about language learning style and preferences
Mid-course self-evaluation
Notes from conferences with peers and teacher

II. Academic work: group projects, individual assignments, tests, quizzes, exercises
Speeches, drafts and final copy
Audio/video tapes demonstrating speaking skills
Evidence of writing skills: Essays and revisions, Journals, Class newsletter
Debate materials
Vocabulary quizzes and corrections
Written comments from peers or instructor about student performance
Favorite learning activities, exercises, or tasks

III. Personal Section

A. Achievements and Honors
Language honors and distinctions
Score reports
Course work from other language classes

B. Personal Interests and Talents
Description of clubs/organizations the student belongs to
Stories and photographs of trips abroad
Creative work: artwork, poems and stories
Articles or books the student has read recently
Favorite inspiring quotations
Favorite songs, books, movies

V. Portfolio Evaluations
Evaluation forms: self-evaluation, peer evaluation, instructor evaluation
Self-evaluative report about own portfolio
Letters to the portfolio creator from peer and instructor about portfolio quality

For each item included in the portfolio, the students write a *cover note* - a brief analysis which explains what they did to complete the work and why they chose to include it in the portfolio. (See handout C, *Guidelines for Writing Cover Notes.*)

4. How do students choose the work to include in their portfolios? When do they make these choices?

A major activity throughout the whole portfolio process involves deciding which work goes into the portfolio and explaining why. The students ask themselves three basic questions when sifting through their work:

What did I do for this work?
What did I learn by doing this work?
What does this work reveal about my language goals, interests or progress?

Students can do much of this 'sifting' work at home, but they also spend class time in small groups looking at work and answering these questions. I walk around the room listening to their discussions, offering advice when they ask. This group work helps the students feel more comfortable with the selection process and also gives them a chance to see the choices their peers were making.

I found that the best time to begin making serious choices for the portfolio was half-way through the semester, after the students had collected about one to two months worth of work. For some students, their decisions were made quickly and early on. For others, this selection process was not complete until a day before the portfolio was due. Some students appreciated being given deadlines so that they didn't procrastinate about the selection task. Four weeks before the final deadline, I scheduled 15-minute conferences with each student and required them to bring a written inventory of their portfolio contents and at least one completed cover note; these meetings were beneficial because they allowed me to see how each student was visualizing his or her portfolio and prompted the students to (nearly) finalize their portfolio choices and to begin working on their cover notes.

At the beginning of the semester, many students felt frustrated that "they had no good work to pick from." To aid their selection task, it was important that students stored all their classwork in one place (a box or file) so that they were able to sort through their growing collection several times over the semester. It also became more imperative that I plan activities that were contextually rich and meaningful for the students. Often, this meant planning lessons where the students worked on group projects, learned to assess their peers, or could reflect on their own learning. I found that if the students engaged in richer and more authentic language learning activities (e.g. a discussion about the purpose of education, a debate on women's role in the military, a personal essay about the best path to achievement), the decisions about their portfolio contents were more deliberative, and as a result, their portfolios were more compelling.

5. How should the portfolio be judged?

It is very important to develop clear criteria for judging the overall quality of the portfolio, and to make sure that these criteria are discussed and understood by the students before they begin finalizing their portfolio. Setting clear criteria is also important so that anyone reviewing the portfolio will be able to interpret the significance of the student's work.

The criteria for judging a portfolio is decided relative to assessment purpose and portfolio contents. In our framework, the final evaluation (see handout D, *Portfolio Evaluation* and handout E, *Guidelines for Evaluating Portfolios*) is comprised of three layers of evaluation:

1. Written feedback evaluating the portfolio in the form of a letter

2. Ratings for each category of portfolio work on a scale of 1 (poor) to 4 (excellent), accompanied by remarks of strong and weak points for each section;

3. Ratings of the overall portfolio quality on a scale of 1 to 4 in four areas, as shown here:

The portfolio should demonstrate:

Growth
The portfolio demonstrates skill improvement, goal attainment, increase in confidence, and/or enhanced self-awareness about oneself as a language learner.

Voice
The portfolio contains evidence of the student's reflective thinking, creativity, opinions, interests and/or future goals.

Accuracy
The portfolio contains work which reveals the students' ability to control English vocabulary and structure and to use English with reduced errors.

Organization
The portfolio should demonstrate the students' attention to detail, thoroughness, neatness and presentation of contents.

Each portfolio is evaluated three times: by the portfolio owner, a peer, and the instructor. The ratings (1-4) are indicators of how well the portfolio documents student work and how well it conveys the individuality of the portfolio creator. The written feedback and ratings are specifically designed to give students feedback about the readers' impressions of their work. No numerical score is given to the portfolios. Completion of the portfolio project is worth 40% of the students' final grade (remaining grade distribution: 10% homework, 10% quizzes, 10% attendance, 30% participation).

6. How did the students react when the portfolio task was initially introduced?

Many of the students were accustomed to assessment tasks that were single timed-performances, so they react to portfolios with mixed feelings: curiosity ('You mean I get to choose the works I like best?', skepticism ('Is this a real test?'), and a bit of anxiety ('Too difficult! I don't have so much time!'). Students became more comfortable with the portfolio concept with on-going discussion of their possible selections and reflection on their learning. They raised a lot of questions about deadlines, page limits, portfolio requirements, and content options. They also wanted to know what reflective thinking is and how to write cover notes.

Students' concerns and questions are usually best addressed through group discussion. Once or twice during the process, set aside 30 minutes for the whole class to write down (anonymously) any question/concern about the portfolio on slips of paper. Collect the slips, read them aloud, and discuss solutions as a class. This group

sharing of concerns and questions can be therapeutic; it reduces anxiety about the portfolio, and helps everyone to feel more in control of the portfolio direction.

The portfolio project can be intimidating, especially since so much is expected of the teacher and students. The students come to class with differing needs and interests so their reactions to the portfolio will vary. For these reasons, setting clear assessment purposes and fostering an environment where students can share ideas and take risks becomes essential in the portfolio process.

7. Once the students are engaged in the portfolio process, what are some of the biggest challenges they face?

Many students need extra time and guidance while completing their portfolios, because their language skills are not yet developed enough to express their introspections in English, or they are not accustomed to thinking about their own learning.

Thinking about one's own learning and performance indeed is serious and challenging work. *What work am I most proud of? What are my goals? How are my language goals changing over time? When do I know I've done good work? What does my portfolio reveal about me and my learning style?* These are difficult questions, particularly for students who have not had many opportunities to form an opinion about their own learning. At first, it may be advisable to ask the students to answer one or two of these questions on a weekly/biweekly basis in learning journals or in conferences with the instructor. (See handout F, *Questions for Reflection.*) As students become more comfortable with answering these types of questions, they can work towards organizing their thoughts into reflective essays or cover notes on their work. Building reflective activities into the curriculum helps to familiarize students with portfolio reflection on an on-going basis. This integration also sends out a message to students that reflective learning has value and purpose.

Another challenge for the students is finding time outside of class to work on their portfolios. A few students commented that the portfolio was "too tiresome" or "too much" for one semester. To keep students (and you) from feeling overwhelmed, consider narrowing the scope of the portfolio to one area of learning (e.g. vocabulary development), so that the portfolio project is more manageable. As mentioned before, scheduling class time to work on portfolios (e.g. 30 minutes a week, or one full class period a month) can be beneficial. It will encourage students to focus on their portfolios, and restore their enthusiasm for the project when they see other students working as well.

8. What kind of portfolio activities can I build into my curriculum schedule?

What constitutes a portfolio activity depends on the needs and the purpose of your portfolio. One activity might simply be to spend the last five minutes of class discussing in groups: *What did I learn today? What interesting ideas did I hear in class today?* More extensive activities are provided at the end of this chapter in a chart entitled *Building Portfolio Tasks in Classroom Instruction.* The chart suggests some portfolio tasks which I felt effectively made use of classroom time: the activities facilitated the planning, organization, and analysis of portfolio ideas, while enabling students to practice their language skills.

BEST COPY AVAILABLE

9. What are some of the biggest challenges the teacher faces?

A great source of frustration for me was the grade dilemma. I felt defeated at times when I realized that at the end of the portfolio evaluation period, there still was the grade to contend with. I gave out many A's the semester I used portfolios; I had heard that there was an 'unspoken opinion' at Japanese universities that teachers who give out too many A's are too easy and ineffective, but I tried to set that 'rumor' aside as I confidently turned in my grade sheets. I knew that my students had done good work. Now as I look back, I am not confident that the number grade I assigned was accurate, fair, or informative, but I am confident that the portfolio provided me with a lot of "assessable information"---information which was recognized as "real" by the students and me.

Portfolio management--balancing resources of time and energy--was another great hurdle. I spent a lot of time contemplating the portfolio framework and responding to the students' classwork and their portfolios. I sometimes worried that the time spent working on the portfolio framework pulled me away from my other regular teaching responsibilities. I was certain, however, that the process would have positive feedback, so I accepted the risks and burden as natural consequences of the entire project. While it is undeniable that portfolio assessment is rigorous, I think the difficulties of portfolio management are less threatening if the teacher works with other faculty to develop assessment purposes and evaluate the portfolios. One way I would improve the portfolio process would be to build in meeting times with other faculty to get input and guidance. This collaboration undoubtedly would help me gain perspective on the portfolio by sharing my experiences with others, and thereby helping me to avoid feeling overwhelmed by the process.

10. What can I do when all the portfolios are completed and I've finished evaluating them?

First of all, celebrate--you and your students have completed a very rich and challenging process. Encourage your students to share their portfolios with their peers, their teachers, and family members.

Share your results and portfolio concerns with other faculty and administration, or present at a conference about its potential in the Japanese university classroom. Perhaps, most importantly, spend some time reflecting on the impact the portfolio has had on your classroom by writing in a journal or chatting with colleagues.

Student Handout A: The Portfolio Process

Teaching Suggestion: Copy this outline for the students as a note-taking sheet when you introduce the portfolio. Or, students can brainstorm about possible activities which may take place during each stage of the process.

The Portfolio Process

Preparation

Gathering and Sifting of Work

Reflection

Construction

Evaluation

Sharing

Student Handout B: Introduction to the Portfolio

Introduction to the Portfolio

What is a portfolio?

A portfolio is an organized collection of your best work in this English course. Your portfolio will show the products and process of your learning. It will also demonstrate your progress and accomplishments in all four skill areas: reading, writing, speaking, listening.

Outline of Portfolio Contents

I. Personal Statements
- A. Reflective Work: Student's beliefs about learning
  - Portfolio Introduction
  - Personal reflective essay (Topic: My Definition of Achievement)
- B. Self-Assessment Documents
  - Letter of self-introduction to the teacher
  - Needs analysis of language goals and interests
  - Notes from conferences with peers and teacher

II. Academic Work
- Group projects, individual assignments, tests, quizzes, exercises

III. Personal Section
- A. Achievements and Honors
- B. Personal Interests and Talents

V. Portfolio Evaluations
- Self-evaluation, peer evaluation, instructor evaluation

GUIDELINES FOR WRITING COVER NOTES

Instructions: For each work you include in your Academic Work and Personal Sections, write a cover note. Please use the following outline to organize your thoughts.

DESCRIPTION OF ASSIGNMENT

First Paragraph What did you do to complete this assignment? What skills did you focus on?

ANALYSIS OF LEARNING

Second Paragraph What did you learn by doing this project? How does this assignment demonstrate that you have improved your skills or increased your knowledge about English? Are you happy with the grade you received? Why or why not? (If there are things you could change about this assignment, mention them there.)

ANALYSIS OF SIGNIFICANCE

Third Paragraph Why did you decide to include this work in your portfolio? What can a reader find out about you based on the quality of this assignment?

Student Handout D: Guidelines for Portfolio Evaluations (2 pages)

GUIDELINES FOR EVALUATION OF PORTFOLIOS

Criteria for Judging Portfolio Quality	
Growth	*The portfolio demonstrates skill improvement, goal attainment, increase in confidence, and/or enhanced self-awareness about oneself as a language learner.*
Voice	*The portfolio contains evidence of the student's reflective thinking, creativity, opinions, interests and/or future goals.*
Accuracy	*The portfolio contains work which reveals the students' ability to control English vocabulary and structure and to use English with reduced errors.*
Organization	*The portfolio should demonstrate the students' attention to detail, thoroughness, neatness and presentation of contents.*

GUIDELINES FOR EVALUATION OF PORTFOLIOS

SELF EVALUATION (Deadline: ____)

1. Write a letter to the portfolio reader, following the guidelines below.
2. Complete the Evaluation Form on your own portfolio.
3. Type or neatly write this letter in ink. (minimum 2 pages, A4, double spaced)

Dear Portfolio Reader,

- Tell the reader something you are especially proud of in this portfolio. What areas are the most impressive in your opinion? Explain why you feel this way.
- If there are areas you think need more work, explain how you would improve or change them.
- Explain to the reader how you feel about your portfolio and how you feel about making it. What will you remember about making a portfolio several days, weeks, years (!!!!) from now? What have you learned about yourself from putting together this portfolio?
- Anything else you wish to say ...

Sincerely, *your name*

PEER EVALUATION (Deadline: ____)

1. Write a letter to the owner of the portfolio. Follow the guidelines below.
2. Complete the Evaluation Form on your peer's portfolio.
3. Type or neatly write this letter in ink. (minimum 2 pages, A4, double spaced)

Dear _____,

- Congratulate your peer on something you particularly enjoyed or found interesting in this portfolio. Try to explain why it had this effect on you.
- Explain the overall effect his or her portfolio had on you. Which parts are you likely to remember several days from now?
- Anything else you wish to say ...

Sincerely, *your name*

Student Handout E: Portfolio Evaluation Form (2 pages)

PORTFOLIO EVALUATION FORM

Portfolio Owner: _____ Evaluated by: _____

Portfolio Contents

Rating Scale: 1 = poor,.... 4 = excellent

Column for comments: ☺ = Impressive areas ☹ = Confusing areas, areas needing improvement

Section	Rating	☺	☹
Portfolio Introduction	1 2 3 4		
Personal Reflective Essay	1 2 3 4		
Academic Works (Listening/speaking)	1 2 3 4		
Academic Works (Reading/Writing)	1 2 3 4		
Personal Section	1 2 3 4		

OVERALL PORTFOLIO QUALITY

Section	Rating	☺	😐
Growth *Knowing more about oneself as a language learner* *Increased confidence* *Improvement in skills* *Setting and meeting goals*	1 2 3 4		
Voice *Expression of one's own reflective thinking, creativity, opinions, interests and future goals*	1 2 3 4		
Accuracy *Ability to control English vocabulary and sentence structure* *Ability to use English with reduced errors*	1 2 3 4		
Organization *Attention to detail, neatness and presentation of contents*	1 2 3 4		

Some Questions for Reflection

What have I learned about myself as a learner in this class?

What makes me feel proudest (or most dissatisfied) regarding my English abilities? Why?

What is the most significant thing to happen to me in this class?

When did I feel most positive about my learning? Most negative? Why?

What would I like my teacher to think about me as a learner?

What do I need to learn the most in this class?

What do I worry about the most as an English learner?

When do I know that I've done good work?

Will I continue to study English after university? What are my goals and purposes?

What interests me the most about studying English?

What is the best way to learn English? to assess English abilities?

***CHART*: Building Portfolio Tasks Into Classroom Instruction**

	TARGETED LEARNING SKILLS				
Sample Portfolio Task	Oral Skills	Writing	Reading	Analy-tical Thinkin g	Collabo -ration; Negoti a-tion
SPEECHES. Students give a 3-5 minute impromptu speech about their personal philosophy about learning.	✔			✔	
BRAINSTORMS. Students learn to brainstorm about various topics related to portfolio contents and process. *Possible Topics:* What is the best way to test English skills? What will I put in my portfolio? How do I define a portfolio? What are the rewards and problems of making a portfolio?	✔			✔	✔
ESSAYS. Students write a personal essay on the definition of achievement. They learn how to peer-read and peer-edit by sharing drafts with their classmates.		✔	✔		✔
CONFERENCES. Students meet with the teacher for a 15-minute discussion about their portfolio contents and questions about the process.	✔			✔	
READING-BASED DISCUSSIONS. Students read an article about conflicts in student-teacher relationships, and then answer comprehension and discussion questions in class. One student in each group is chosen to be the leader of the group discussion.	✔		✔	✔	✔

REFERENCES

Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey-Bass.

Calfee, R. (1994). Ahead to the past: Assessing student achievement in writing. Berkeley: National Center for the Study of Writing.

Davidson, B. (1995). Critical thinking education faces the challenge of Japan. [http:/www.shss.montclair.edu/inquiry/spr95/davidson.html]

Gottlieb, M. (1995). Nurturing student learning through portfolios. *TESOL Journal, 5* (1), 12-14.

Harris, M. (1988). *Women and teaching*. NJ: Paulist Press.

Huerta-Macías, A. 1995. Alternative assessment: Responses to commonly asked questions. *TESOL Journal* (autumn): 8-11.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Paulson, L. (1993). Bridging the gap: Applying Deming's "quality by design" to portfolios. Paper presented at the Portfolio Conference in the Heartland. Des Moines, Iowa. ED 378 148.
SP 035 666.

Sheingold, Heller, & Paululokonis. (1995). Actively seeking evidence: Teaching change through assessment development. NJ: ETS, Center for Performance Assessment.

Sweet, D. (1993). Student portfolios: Administrative uses. In *Educational Research Consumer Guide 9*. Washington, D.C.: U.S. Department of Education. [http://inet.ed.gov/pubs/OR/Consumer Guides/classuse.html]

Tierney, R.J., Carter, M.A., & Desai, L.E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon.

Valdez-Pierce, L. (1996). Assessment and instruction: Separate and unequal? *TESOL Matters, 6* (1): 9.

Chapter 4

Incorporating Alternative Assessment in the Teaching of EFL Writing: Responding to Multiple Drafts

SHALLE LEEMING

PARADIGM SHIFTS: FROM PRODUCT TO PROCESS

The term *paradigm shift* was first used by Thomas Kuhn (1963) to describe the process by which major changes come about in scientific fields. He hypothesized that revolutions in science occur when old intellectual systems no longer fulfill the needs of new problems.

This term has more recently been used by some language and writing teaching professionals to describe the shift that has occurred in the approach to the teaching of both ESL/EFL and first language writing (Hairston, 1982). In the last twenty years, the most noticeable shift has been in the movement away from a more traditional paradigm, which approaches the teaching of writing through an analysis of discourse. In this approach, students primarily study organizational patterns, grammar, syntax (sentence structure and sentence boundaries, etc.) and mechanics (spelling, punctuation, etc.) by looking at models of writing and trying to emulate them. This approach is sometimes referred to as the "product approach" because instruction focuses on what an ideal final product should look like.

This can be contrasted with a recent move towards a more process-centered approach to the teaching of writing which seeks to go beyond the study of models and a focus on the final product by spending a large amount of time on teaching students the process writers go through to create meaning.

Research into the composing processes of native and non-native writers in English has revealed that creating effective writing is a complex endeavor (Zamel, 1983; Murray, 1978). Writing effectively, the studies suggest, is not purely a matter of form and language. It is not simply a matter of mastering verb tense or memorizing rhetorical forms; it is a process that requires a writer, among other things, to gather thoughts, look for relationships between ideas, construct arguments, and address a reader's questions, objections, and expectations. It is from this understanding of the complexity of the writing process, combined with insights into learning styles and cognitive learning theory, that language teachers have begun questioning and experimenting with the way they assess and teach writing in a second or foreign language.

TENETS OF ALTERNATIVE ASSESSMENT

Indeed, it is this questioning of and dissatisfaction with existing standardized tests that led to the alternative assessment movement in education, a movement dedicated to developing new styles of systematic assessment in education. Some of the central tenets of alternative assessment practice, as paraphrased from Herman, Aschbacher, and Winters, 1992, are:

- Assessment should examine the processes as well as the products of learning.

- Assessment should go beyond merely testing routine and discrete skills to tap higher- level thinking and problem solving skills.

- Assessment should try to integrate assessment methodologies with instructional outcomes and curriculum content.

- Specific criteria and standards for judging students' performance should be set.

- An integrated and active view of learning requires the assessment of holistic and complex performance

- Assessment systems that provide the most comprehensive feedback on student growth include multiple measures taken over time.

In this paper, I would like to consider some of these tenets and how they relate to the use of multiple drafts in the evaluation of writing. Then I would like to discuss some guidelines for feedback at various stages in the composing process. I realize that teachers hold various beliefs about the nature of teaching, learning, and education. I present these ideas as a possible alternative or supplemental measure for university language teachers who are interested in exploring these new perspectives in the teaching of writing.

CURRENT PRACTICES AT KGU

In the Kwansei Gakuin Assessment Survey (which is discussed in Chapter 7), 16% of the respondents who saw their classes as primarily focused on writing explained evaluation procedures that assessed more than one version of a writing assignment. Though it is possible that some respondents give students informal feedback but did not detail that practice on the survey, 41% stated that they only evaluated students' expository writing ability through their performance on the final exam. This suggests that the majority of university language teachers at Kwansei Gakuin evaluate writing ability by the assessment of a single draft.

WHY ASSESS MULTIPLE DRAFTS?

Certainly single draft assignments and in-class essays still have a place in evaluation, but there has been a movement away from single draft assessment of writing, especially for more complex writing assignments such as composition, and especially in its use as the primary or exclusive evaluation of student progress.

One of the reasons for the emergence of multiple draft writing assessments--assessments in which a student receives feedback from a teacher or peer and then revises, sometimes as many as two or three times--is related to the alternative assessment philosophy that evaluation practices must pay more attention to the learning processes of students (Herman, Aschbacker, & Winters, 1992).

Research on the impact of teacher's comments on student writing suggest that feedback during the process of writing may have a greater impact on student learning. A study by Burkland and Grimm (1986) revealed that students are more likely to reject or ignore feedback on works they consider to be dead: final drafts with grades. This suggests that feedback on writing can be more effective if it comes at a time when students are more likely to benefit from it, when they are still struggling to make

meaning. By intervening during the process, teachers de-emphasize the final product, and, instead, emphasize the learning process of the students.

Another justification for assessing multiple drafts is that it more closely mirrors and models the writing process of real writers. Research into the composing processes of writers has demonstrated that writing is a task of some complexity (Perl, 1980; Zamel, 1985). These studies reveal that the process of making meaning is not linear; it is a recursive process that involves finding, focusing, clarifying, and organizing ideas. If we observe and interact with students as they negotiate this complex process by helping them see their strengths and weaknesses as a writer, and by setting clear priorities for revision, we may be able to have a stronger impact on the development of the writer. Proponents of this approach to assessment, who place a high emphasis on the development of critical thinking skills and problem solving skills, see this kind of skill building--and, consequently, this type of assessment--as a necessary component of teaching writing in a foreign or second language.

The use of multiple draft assessment is also closely linked to concepts of cognitive learning theory. If we accept that our students vary in terms of learning style, developmental pace and type of intelligence, we must also look for ways we can determine what each student's strengths and weaknesses are and try to accommodate them. The additional information gleaned by the assessment of multiple drafts, and the closer interaction of the teacher and the student, may provide more of an opportunity for teachers to observe and address individual learning needs.

The proponents of alternative assessment techniques believe that performance-based activities alone do not necessarily give us the full picture of students' abilities. Multiple measures over time, they believe, may tell us more about a student's abilities. Responding to one or more intermediate drafts may be a better way to learn about and respond to our students' needs.

RESPONDING TO STUDENT WORK

The following are some suggestions for responding to student work:

- **In the early stages, focus on content.**

Students need to know that the content of their message is of primary importance. We can help them establish this priority by focusing our comments on content concerns. Research into the effect teachers' responses have on the revision of texts (Zamel, 1985) found that when teachers made comments about both content and grammar, the more important content concerns were often ignored by the students. Part of the reason may be that these comments often contradicted each other. Zamel found that teachers would sometimes question the relevance of a particular section, and then make extensive comments asking for more details or commenting on surface level errors. When faced with mixed messages, students often opted for the easier route of correcting the "errors" but ignoring the larger and more important problems related to meaning.

Since the content of the message is the overriding concern, and because surface level errors may be edited out in the final draft, reserve giving feedback on surface level errors--such as grammar and punctuation--until a later draft.

- **When commenting on errors in later drafts, try to diagnose the source of the errors.**

Let the student know why he/she is making the error if you can determine the cause. Error pattern recognition is more valuable than the correction of individual

errors. If a student is unaware why she/he has made an error, identification or correction alone will have little impact on improving grammatical accuracy. By assisting students with the identification of their grammatical problems, we make those grammatical features more salient, thereby making it more likely that students will assimilate error feedback. With high frequency errors, help them to see that 10 errors on the page may really be only one error.

- **When commenting on errors in later drafts, prioritize errors and limit feedback to only the most important ones.**

One way of prioritizing sentence level errors is to identify them as "global" or "local" errors. Global errors are errors that usually affect the reader's meaning because they tend to affect more than just a small part of a sentence. Local errors are errors that usually affect a small part of a sentence, and while they are distracting, they generally do not impair understanding (Burt & Kiparsky, 1972). Since global errors are more likely to impair meaning, they should be addressed first. Errors with verb tense, word order, modals and verb form are examples of global errors. Errors with subject/verb agreement, word form, and articles are examples of less serious, local errors. However, even if they are less serious, you may want to address them if they are high frequency errors.

Research into the affects of error feedback on student learning suggests that EFL writers can absorb only a small amount of corrective feedback (Robb, Ross & Shortreed, 1986). Consequently, focusing on a few errors may have a larger impact on learning.

- **Instead of correcting errors, determine a system for marking them and have students revise.**

Cognitive theory tells us that when students are engaged in a problem solving activity, long term learning is promoted. Here are some techniques used by writing teachers which cultivate such student generated, long term learning.

1. Put a check mark in the margin of a line that contains an error that you believe the student should be responsible for.

2. Write an appropriate question in the margin to inductively lead the student to find an error you believe they should be responsible for.

3. Correct errors in one section as a "metaphor" for the paper.

4. Don't mark anything, but write an endnote about a few problems the student needs to concentrate on in a revision.

CONCLUSION

Though three of the four suggestions I have included focus on error correction, this is not because error correction is most profitably the focus of writing feedback: exactly the opposite. Too much attention paid to error correction too early on can actually have a negative effect on the writing process, as students may focus on form over content or be overwhelmed with the number of errors in their papers to the point where they become blocked or--worse yet--come to dislike writing.

The feedback we give can dramatically affect the way our students feel about what they are doing. In that it takes great courage to begin to put words down on a blank page, especially in a second language, we can honor that courage by responding in ways that recognize both *what* the author says as well as *how* she/he says it. By giving our students careful, useful feedback on drafts of student work, we validate a paradigm shift not away from product but toward process *plus* product. I believe that this alternative to more traditional modes allows our students more room to see themselves as *writers*, not only as students with a writing assignment due in the morning.

REFERENCES

Burkland, J.,& Grimm, N. (1986). Motivating through responding. *Journal of Teaching Writing, 5* (2), 237-247.

Burt, M.K. & Kiparsky, C. (1972). *The gooficon: A repair manual for English.* Rowley MA: Newbury House.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication, 33 (1),* 76-88.

Herman, J.L., Aschbacher, P., & Winters, L. (1992). *A Practical Guide to Alternative Assessment.* Alexandria, VA: Association for Supervision and Curriculum Development

Kuhn, Thomas. (1963). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Murray, D. (1978). Internal revision: A process of discovery. In C. Cooper and L. Odell (Eds.), *Research on composing* (p. 85-103). Urbana IL: National Council of Teachers of English

Perl, S. (1980). Understanding composing. *College Composition and Communication, 31(4),* 363-369.

Raimes, R. (1983). Tradition and revolution in ESL teaching. *TESOL Quarterly, 17* (4),535-552.

Raimes, R. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly, 19* (2), 229-258.

Robb, T. N., Ross, S. & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20* (1), 83-95.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19,*(1), 79-97.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17* (2), 165-187.

Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly, 16* (2),195-210.

Chapter 5

Student Self-Assessment: A Viable Alternative

JOHN CHAPMAN-RIENSTRA

INTRODUCTION

This chapter will focus on using self-assessment as a means of evaluating students' classroom performance and as a means of empowering students--getting them to take more of an active role in the language learning process. First, I will discuss some background as to how I came to use self-assessment procedures in my classes, and present a rationale for implementing self-assessment practices. Then, I will describe a number of classroom practices that I have used, including step-by-step procedures and a review of how well they worked. The purpose of this chapter is to help teachers expand the forms of assessment they use in the classroom so they might more effectively serve their students.

APOLOGIA FOR SELF-ASSESSMENT

As a teacher of ESOL, one of the biggest problems I face each term is how to fairly assess my students' performances--in other words, how to give them a fair grade. After all, on what criteria am I grading them? What are the "standards" I have for assessing my students' performance? Do my evaluations of their performances truly reflect the level or depth of learning they have achieved?

I have always been dissatisfied with arbitrarily assigning grades to students, particularly if they have had no input in the decision-making process. Even my most sophisticated evaluation techniques still end up being a shot in the dark, and usually my grades end up being based on a student's effort (participation) and the scores they got on my tests, rather than on the extent of language learning that actually took place. In other words, their evaluation is largely just my opinion of how much they have learned, how well they have learned it, and how hard they tried. Is this type of grading system really satisfying to students? More importantly, is it *fair*?

Let's look at this issue from another angle: What are your students' attitudes toward learning English? Do they seem motivated? Excited? Do they have goals? Do they take initiative? Are they diligent? In many English classrooms in Japan, unfortunately, the answer on all counts (except perhaps the last) is "no." In secondary school, the students' overriding goal--and motivation--is, of course, passing the entrance exams. But they do not seem to have any other learning goals. In addition, too often, college students do not know the reasons why they are studying English, other than to fulfill a requirement. Or, if they do have a goal, it is an amorphous one: "to learn more English" (whatever that means).

Certainly, students want to learn English, but if they do not have clear goals in mind, they are simply working for the intrinsic benefit of whatever "knowledge" we can pass along to them. Even worse, if they *do* have clear goals, they are almost never asked about them, so it is just good fortune if theirs matches those of the instructor. More often than not, it doesn't. Instead, students end up doing only the work that is required of them so they can receive a certain passing grade in the course, leaving them

dissatisfied with the course or unsure about what they have really learned--and why. This begs the question, what *have* the students really learned in these classes? Are they themselves even aware of what they have learned or how they have learned it? Do they even think about such things? I think they should.

Furthermore, as mentioned above, many teachers are not in the habit of asking their students to establish their own learning goals or develop procedures for learning in a conventional classroom. Yet, to my mind, this is a vital component of any healthy learning environment. After all, the classes we teach are really for the students' benefit; so why should they not have a say in establishing the objectives--if not the criteria--for learning? This is not to say that the teacher should simply hand over control of the class to the students and let them run it. Instead, what I mean is that students often *do* have goals for their language learning, which we teachers often overlook or ignore. To do so, though, is to rob the students of taking a vital interest in the learning environment; this is also the case with the way students acquire new skills and how well they learn them.

I am a firm believer in the philosophy that students need to take a vital personal interest (often called "ownership") in a language if they are going to learn it well or benefit from it. I also believe that students need to take greater responsibility for their own learning. This is even more strongly the case with language education in Japan, particularly English--the great albatross of the entrance exam system--which is usually taught like Latin: a dead language that no one speaks; students learn its grammar only, so they can pass the entrance exams. All English teachers in Japan are aware, I am sure, of the drawbacks of the English education system here, and the snail's pace in reforming it. But we should be able to do something about these drawbacks at the university level, where we have more freedom in what and how we teach. English is not Latin, so we should avoid teaching it as such.

We also know the result of the current English education system in Japan: it produces students who are incapable of communicating effectively in English. And students are not happy about this at all. During my five years of teaching in Japan, in fact, I have conducted an informal and (admittedly) unscientific survey of all my Japanese high school and college students regarding their level of satisfaction with their English education. Every student I have ever asked--and this now numbers almost 500--without exception, has expressed dissatisfaction and/or displeasure with the English education system in Japan.

There are many reasons for this, but here are a few contributing factors, which have been told to me by students. First, students usually do not know why they are studying the material they are presented--except for the vaguely-stated reason, "it will help them pass the entrance test." In addition, they have not chosen the material, nor have they had any say in how it is taught to them or--most importantly--how they are evaluated.

Although all of these factors deserve close attention, I want to focus here on the methods of evaluation. Typically, in Japan, English evaluations take the form of objective-criteria (usually multiple-choice) tests. But, English is not math. We therefore should not be attempting to measure skill in English the same way we measure skill in math. At the risk of sounding didactic, I also believe objective-criteria tests do not accurately reflect what the students know about the language, since these tests measure only a part of what the students are supposed to have learned. As Hakner and Cutolo (1997) point out, "A single test score cannot accurately assess students' language skills and abilities" (p. 10). But I think objective-criteria tests are particularly unreliable assessment tools when used alone because they do not require learners to *use* the language, which is the basic test of the knowledge of any language: can the learners *use* it, and if so, to what extent?

Let me return to the central problem of assessment, however; which is: What have the students *really* learned, and how can an instructor measure this accurately? First, by "learned" I do not simply mean "memorized"; I mean "acquired"--what the

students have been able to incorporate into their long-term memory and inner consciousness. In other words, what elements of language are students able to use and not just regurgitate in drills or guess correctly on a test?

Secondly, the real problem at hand: how do we measure this learning? I will be the first to admit that I am not a psychic; I cannot pretend to be able to delve into the minds of my students and cull this information from them. I will also be the first to admit that conventional methods of assessment, such as multiple-choice and fill-in-the-blank tests--which I have and still do use--do not always give me an accurate picture of what the student really knows and is able to use. So, how can teachers find out if students have reached the learning objectives that have been set out? One solution is to ask the students themselves. In other words, leave the assessment of the students' learning in their own hands.

RATIONALE FOR SELF-ASSESSMENT

These thought processes led me to start thinking of ways in which I could let my students have more control--and therefore ownership--over their own English learning in my classes. The answer I came up with was to begin incorporating student self-assessment into my course curriculum. Part of the reasoning for this is based on the fact that I am a proponent of a learner-centered curriculum, and especially experiential learning, which is characterized by (among other things): interactive communication in English, task-oriented learning, a focus on both the language and the learning process itself, and integrating the students' own personal experiences into the learning process. I consider this a crucial element in my classroom instruction. And, as stated above, I also believe that students need to take greater responsibility for their own learning. The best way to accomplish this, to my mind, is to give them choices about what they want to learn and require them to monitor their own progress.

Again, although I would like to touch upon giving students learning choices, my focus here is on having students assess themselves. Why self-assessment? First, it helps to create a more learner-centered classroom environment, which in turn helps students develop a greater awareness "of their own role as active agents within the learning process" (Nunan, 1988, as cited in McNamara and Deane, 1995). It can also help students identify their own strengths and weaknesses in English and develop more effective learning strategies (McNamara and Deane, 1995).

Second, research has shown that self-assessment stimulates students to a greater commitment to the learning process and a deeper understanding of learning strategies. John Cowan (1988, p. 206: referring to Boyd and Cowan, 1985) writes that the students' self-monitoring of their learning, "which entails thorough identification and comparison of performance with criteria, stimulates a questioning approach to both elements in the comparison." Consequently, "Self-assessment had a powerful role in changing the quality of learning and of learning aspirations for many of the students in the class" (p. 204).

Third, as Huerta-Macías (1995) concludes, there is no reason to consider alternative assessment techniques, such as self-assessment, any less objective or accurate than traditional testing. She argues that standard objective-criteria tests are not really "objective"; they are just subjective in different ways (p. 10).

Finally, self-assessment can become a good complement to traditional evaluation techniques. McNamara and Deane (1995) point out that using both types of assessment techniques can help teachers to create a more complete picture of their students' abilities, effort and progress. "More importantly," they add, "students have a greater voice in their language learning process " (p. 21).

Perhaps an analogy here would make my approach to self-assessment clearer. In some ways, learning English is an art because of the creative processes involved. In order to become a painter, for example, one must first study the basics of the art--its forms, techniques and styles--and study the examples of some masters. This would be

analogous to an English student learning the basics of English--grammar, language structure and vocabulary--and studying the language with fluent speakers. Once a student has learned the basic principles, however, they need chances to practice in order to master their technique (and thus, their art). Once they have reached this level of development, it is very possible for aspiring art students to start exploring on their own, to develop their own expressive style and to set goals for themselves and ways to achieve those goals. They also need to be able to judge the quality of their own work, or they may never learn to grow on their own and master their art. The same can be said for students of English: once they know the basics of the language, they are fully capable of setting learning goals for themselves, and are at the very least able to judge the quality of the language they produce. True, this skill requires a certain level of maturity, but by university, students most certainly have developed the basic language skills and the cognitive skills necessary to be able to assess their own work. Cowan (1988) concurs: "Undergraduates showed themselves able to cope with demanding objectives, self-direction and self-assessment" (p. 206).

It is also quite possible to look at self-assessment in a more objective, scientific framework. If we consider learning a language to also be a scientific process of learning rules and applying them to see if we know how they work, we can compare the work of a student to that of a scientist. For example, before beginning a new project, a good scientist always sets an objective, and plans a way to reach it. A scientist typically asks: "What objective do I want to accomplish next; what instruments will I need to use; what is the most effective experimental procedure to implement; and how will I measure the success or failure of the experiment?" The same can be said for learning language--in this case, English. A good student needs to ask: "What skill do I want to acquire next; what tools do I need to acquire this skill; what is the most effective way (procedure) for me to acquire it; and what criteria will I use to measure how well I have learned it?"

If scientists haphazardly went about conducting experiments with no clear goals in mind other than "this is what I was told to do," they would not gain an under-standing of the process or be able to interpret the results, rendering the entire process--and any resulting knowledge--meaningless. Neither would the scientists benefit by doing the work themselves and then having someone outside the project evaluate the results. True, the scientists could still learn something from an instructor-directed experiment, in which the teacher told the scientists how clear their results were or how well they had completed the work, but the process of self-discovery and growth on the part of the scientists would be greatly reduced, in my opinion.

This is the key, I think, to student self-assessment: self-actualization, not simply accurate imitation. In other words, students need to know what the language learning objectives are, be aware of how these new skills are acquired, and understand how well they can use them. Of course, I can also accomplish this by telling students how well they use certain language skills myself--which is in effect what we all do with conventional teacher-authored assessment. But in the long run, I believe students benefit more from the self-awareness of the learning process and the self-development they procure by taking responsibility for their own learning and the assessment of it. Much recent research also supports this claim. (For a list of further sources, see McNamara and Deane, 1995, p. 21.)

As a final point, however, I would like to state that the process of incorporating self-assessment into my curriculum has been--and still is--gradual; I have not completely abandoned traditional assessment techniques, for several reasons. First, I think that students need time to get used to the idea of self-assessment and the responsibility of evaluating themselves. If you plan to introduce any self-assessment strategies into your classes, I would recommend doing it gradually, so that students can adjust to what will be a radically different style of learning for them. Consider this a transition phase. Students will not know what is expected of them at first, which means

that you must provide them with enough structure within which to work and detailed models to follow so they can succeed.

This also means you need to know what you are doing, which is the second point: you as an instructor will also need time to get used to a new style of teaching. It is a hands-off style of teaching which is student-centered, not teacher-centered, so it involves a great deal of facilitation instead of authoritarian rule. It also entails letting the students learn how to learn and how to judge their own learning. Employing self-assessment strategies in the classroom has definitely been a struggle for me. My experience with this process over the past two years has been similar to that of John Cowan (1988), who wrote:

> I still find it difficult to plan effective facilitation of learning in which I should not direct or otherwise influence the learner. And I find it even more difficult to help learners develop the ability to judge their own learning, without conveying something of my own values in the process (p. 194).

Again, let me iterate that the purpose of this chapter is to help teachers expand the forms of assessment they use; it is not intended to imply that self-assessment is a panacea for English education. Nor am I suggesting teachers altogether eliminate objective-criteria assessment strategies, such as multiple-choice tests. As mentioned before, I have and still do use these types of assessment methods in my classes. The point I am trying to make is that self-assessment is a viable and effective alternative to traditional assessment strategies and, as mentioned above, it can be a good complement to them.

Hopefully, the examples which follow will provide you with some practical ways you can employ some self-assessment strategies in your English classroom, in a way which will help your students take greater ownership of their language learning and make a deeper commitment to it, which in turn will facilitate a deeper understanding of English.

EXAMPLES (CLASSROOM APPLICATION)

Several examples of self-assessment have been presented already in other chapters in this book; here, I would like to focus on some examples of micro- assessment, by which I mean assessment of an individual activity, and macro-assessment, by which I mean overall assessment of a long-term activity, such as portfolios (see chapter 3).

1. Self-Assessed Oral Vocabulary Quizzes

One aspect of assessing vocabulary which has always been problematic for me is whether or not what I am testing students on actually reflects what vocabulary they really know. In an intensive English course, such as the Intermediate course offered by the Language Center, which meets three times per week, students typically encounter around 30-50 new vocabulary words per week (depending on students' prior vocabulary knowledge). Quizzing students on all these words, and then grading the quizzes, would be a daunting and time-consuming task, especially for large classes. So, when I have employed standard fill-in-the-blank and multiple-choice tests, I have ended up quizzing students on roughly 30-40% of these words only, at best. This seemed to me to be a very inaccurate reflection of the extent of the knowledge students actually have of the new vocabulary; for example, two students might know 80% of the new vocabulary, but on a quiz, one might score 90 and the other 70, depending on the vocabulary they know which is included on the quiz.

There are several solutions to this problem. However, several years ago, I decided to try a less conventional approach: student-assessed oral vocabulary quizzes. This idea came from an article by Tim Murphey in the Winter 1994/1995 *TESOL Journal* ,

from which I borrowed heavily. Murphey summed up his rationale for using this technique this way: "Putting students more interactively in the center of creating and administering the tests reduces anxiety as students feel more in control and actually have fun with the tests" (p. 12).

Background

During the semester, each student in my class compiles a list of unknown words, phrases and idioms in a vocabulary log, which is kept in the student's notebook. Although I assign units from our vocabulary text, students are responsible for making the log themselves and also incorporating any new vocabulary which they compile from other class activities. Students are required to log a minimum of 10 words per week over the semester, and review their log periodically. Several times during the semester (how often to give a quiz, though, is up to the teacher), I give the students the *Vocab List* sheets (see Appendix A1), which they use as a study guide for the oral quiz. They then fill out this sheet with 20 words or phrases which they have not mastered, writing a definition and an example sentence for each one.[1]

Another option--one which has been used by Murphey and myself--is to have students submit a list of words or expressions they have not mastered, which you then use to construct a class vocabulary list handout for the students to use to study for the quiz. Either way, the value of having students choose the material for the quiz is that the students clearly know on what they will be tested. This in turn helps reduce student anxiety about the tests (Murphey, 1995), and helps them perform better. As Murphey (1995) writes, "For the test to be beneficial for my students, I prefer to make the material on any test clearly known to all students. The easiest way to do this is to let them make it" (p. 16).

The procedure

For the quiz itself, students work together in pairs. I have noticed that the chances of cheating occurring are reduced if you assign each student a partner, preferably one they do not know well. This lessens the probability of friends evaluating each other and "fudging" the scores (after all, most people are reluctant to give their friends a failing score). The procedure for the quiz is listed on Appendix A2. However, I have found that I still need about 15-20 minutes to explain the procedure to the students, and walk them through a few examples. In addition, if the students have a variety of proficiency levels, my experience has also found that the test seems to work best if students with similar skill levels are paired together, although this is not a necessity. The reason for putting students at similar levels together is mainly for their own comfort; students have written in their journals to me that they feel less comfortable and confident when they are working with a student who has an obviously higher or lower skill level.

Once students have their partners, they find a place to take the quiz apart from everyone else. This means that students can leave the classroom if they want and go outside or find another place in the building, as long as they are not disturbing other classes. They may just sit somewhere together, or walk around. There are two reasons for this tactic. First, being outside the classroom may make students feel more positive.

[1] It is important for students to write an original sentence for each entry, as simply copying a sample sentence from the dictionary would not encourage acquisition of the word into long-term memory. I tell my students to write an example of how *they* would use the item in a sentence, in a way which indicates they understand what the item means and how it is used.

Also, as Murphey (1995) points out, "other classmates do not hear what they say and they don't feel like the teacher is standing over them; [so] they are more relaxed and can access more of what they know" (p. 14). Second, being able to move around benefits the various learning styles your students have, particularly those who are kinesthetic learners, who dislike being confined to a chair and thus find it easier to walk and talk. However, if your class situation does not provide you with the type of freedom to allow students to leave the classroom, you can still facilitate a more relaxed atmosphere by playing some soft background music while students are giving each other the quiz.

During the quiz, students evaluate each other's performance and negotiate their score with their partner. The value in this process is that students learn "the most important evaluation is ultimately what they think of themselves and that they are responsible for their own learning and judging themselves" (Murphey, 1995, p. 14). For a more detailed description of the scoring procedure, see Appendix A3.

After the quiz, the students hand in their Mark Sheets and their Vocab Lists for me to check. I then go over the Vocab Lists to see if there are any global errors in the definitions or example sentences. But, even if there are serious errors in the entries, I do not alter the quiz score; I always accept the students' scores at face value. It is important not to alter the students' scores because, as John Cowan (1988) writes, "It is your trust in their ability to assess themselves which prompts them to be responsible" (p. 209). So, my review of their Vocab Lists is included in a separate score, as part of their overall daily work evaluation.

Comments

Has this type of quiz been successful? Absolutely. Many students have written to me that this type of quiz is more enjoyable and less stressful. Are the scores that students assess themselves honest and accurate? By all accounts, yes. First, as for accuracy, given the specific parameters of the test procedure, there is little room for misunderstanding the assessment criteria. To check for scoring accuracy, in fact, for each oral quiz I select one or two students to be my partner[s]; I take the quiz with them (I sometimes even make my own list!), and negotiate their scores with them. Almost always, the students' own assessment of their performance has been the same--or lower!--than my assessment. And comparing scores across the semester, the quiz score of the students who have been my partners are consistent with their other scores, which means that students are not scoring themselves strictly only when they are taking the quiz with me. In fact, David Gardner (1996) reports that other researchers have found "a relatively high correlation between peer assessments [of oral tests] and their own expert teacher assessments of each candidate" (Miller and Ng, 1996, as cited in Gardner, 1996, p. 19).

As for honesty, of course, there is nothing preventing students from cheating. This is the real caveat of this type of test: the teacher must give up control of the assessment process and trust the students to evaluate themselves fairly and honestly. Some students may cheat. But, cheating occurs on conventional tests, too. As Murphey (1995) sums up: "We can either spend our time being police, trying to control and measure students completely, or be teachers and offer them opportunities for learning. I think our energy is better spent in the latter" (p. 16).

In general, my experience has been that students tend to do better (i.e., score higher) on this type of quiz than on "conventional" quizzes. This is not because they inflate their scores--in fact, I have had several students fail themselves--but because they know exactly what they will be tested on, and what procedure is involved. Since they control what they are tested on and know what criteria they must meet to succeed, I believe they tend to study much more effectively.

This is not to say that this type of test is flawless. As an evaluation tool, in fact, this type of quiz may not be any more accurate than traditional objective-criteria tests. The difference, really, is the function of the oral quiz. As Murphey (1995) states, "The main purpose of [self-assessed oral] tests is not evaluation but rather stimulation of effective learning processes that can later be used by learners to help them learn whatever they want" (p. 13).

One important advantage of this type of test, for example, is that it is beneficial to all your students, even if there is a large disparity of skill levels among the students in your class. Murphey (1995) writes that this type of assessment technique is "open ended enough so that advanced students can still get something from it and lower level students can still learn and get a feeling of success so they will want to continue learning and studying" (p. 12).

Another advantage of this type of test is the amount of actual oral language which the students must produce. My students have often told me that they want more chances to speak English in our class. The oral quiz is an ideal way for them to get more practice, in a way which is conducive to language acquisition. In the end, these are goals for which all teachers are striving: to help students help themselves and make them want to continue learning.

2. Reading Discussion Groups (a.k.a. Seminars or Seminar Discussion Groups)

Another activity which I have been using for student self-assessment--and to give students more speaking practice--is the Reading Discussion Group (RDG). The RDG is a student-led small group discussion activity, much like a student-led seminar. In it, one student per group acts as a discussion leader and guides the other members through a review of a short newspaper or magazine article and then a formal discussion about questions related to the topic, with students freely exchanging ideas and opinions. It works best with intermediate and advanced students, although it can work with lower-level students if appropriate reading materials can be found for them (see details below). The reading can even be foregone in favor of simply having a formal discussion on a given topic.

Background

The two classes in which I first employed the RDG had between 20-30 students. So, I divided the class into groups of around seven or eight, at first. However, even in a group that size, there were still inevitably several quiet students who would remain wallflowers and not participate actively in the discussion. The next semester, I reduced the size of the groups to six, which worked fairly well but still allowed a few students to sink into the woodwork. This past year, I have limited the size of the groups to five, which has been, so far, the most successful; it is extremely difficult for even the shyest of students to disappear in a group that small, and often the smaller size of the group encourages shy students to come out of their shells and participate. It has been my experience that Japanese students, as group-oriented as they are, seem to function better in smaller--hence less intimidating--groups.

The issue of how to divide the students into groups, however--randomly or by proficiency level--still vexes me. In the first two years of the RDGs, I tried to employ both methods, but they never worked to everyone's satisfaction. For example, in one class in which the students were divided by skill level, some of the lower-level students said that although they felt a little more comfortable in their group, they did not feel as if their English improved as much because there was no higher-level student in their group to challenge their skills. On the other hand, in another class in which the

students were divided randomly, the highest-level students did not feel as challenged because they were all in groups with lower-level students.

Finally, this year, I decided to simply let my students decide how they wanted to be split up. To my surprise, they chose to be split up randomly (I did it according to birth dates, although I did make sure that there was a good cross-representation of all proficiency levels in each group). This past semester, this system seemed to work out well for all my students. My conclusion, therefore, is that the teacher and students should work together to decide which grouping system would be more appropriate for their class.

The setup

Before my students conduct a RDG, it is important that they build some schema--background knowledge and vocabulary--for the chosen topic. I have found that the RDG activity works best if it is used in conjunction with a theme that is being worked on already in class. This theme can come from a text or an independent source, such as TV or films. Typically, at the beginning of a thematic unit, my students watch a short 3-5 minute clip from a film or TV show, listen to a radio broadcast and/or or read a short news article on a certain topic. Some topics which have worked well are: bullying (*ijime*), working women, gay marriages, divorce, public smoking, HIV/AIDS, capital punishment, cram schools and the entrance exam system. There are many others, including simpler ones which are more accessible to lower-level students, such as foreign food and travel. Teachers have many possibilities for topics, depending on students' interests and majors. The point is, it is important to try and select topics which will appeal to your students' interests. I have made the mistake of assuming the topic I choose will be of interest to my students, when in fact it is not. For example, Science Department students do not necessarily want to discuss science-related topics.

I often try to have an informal discussion activity about a particular point in connection with the topic, and I also try to include some related vocabulary as homework or in-class work. Sometimes, with a broader or more popular topic, my students will work on a longer project, such as a small group presentation. All of these activities aim to build students' background knowledge so that they feel more comfortable discussing the topic in an extended formal discussion.

Finally, and most importantly, before you even assign a topic for the first RDG, it is vital that you model every aspect of the RDG for your students first. I took one full 90-minute class to go over all the RDG materials (see Appendix B1-B6), including explaining the RDG procedure, reviewing some tips on expressing agreement and disagreement, and modeling a discussion for the students. I also had the students fill out Self-Evaluations after the Model RDG so they could see on what criteria they would be assessing themselves for the "real" RDGs (see Appendix B7-B8). This helps tremendously in making the student-led RDGs proceed smoothly.

The procedure

Class #1: Assign the readings for the RDG and give the Discussion Outline and Discussion Leader's Checklist to the designated group leaders. With lower-level students--and with students who are unfamiliar with the seminar format--it is more helpful to assign the readings to the leaders. In fact, I try to have several different readings on the same topic for the leaders to choose from, but this is not always possible. For more advanced students, or those who have led a RDG before, you can let the students research the topic and find the readings themselves, although I have always

told my students they can come see me if they have trouble finding one, as I have a large collection of news articles on file.[2]

The group leaders are responsible for perusing the article, summarizing its main points and learning important vocabulary items so that they can in turn explain them to the other group members. They must also write a list of questions to use to help them lead the discussion. I require the leaders to construct two types of questions: information questions, based on material in the article, to check that all the group members have read the article and understand it; and thought questions, which ask the group members for their ideas and opinions about the topic. I have found that 4-6 questions (2-3 of each type) are enough to easily sustain a 20-30 minute discussion, although this will again depend on how good the questions are and how outgoing your students are.

All of this is assigned as homework for the leaders for the next class, although the actual time frame should be about 3-7 days. It is also helpful, if you have lower-level or inexperienced students, to meet briefly with each of the group leaders before the next class to go over their Discussion Outlines, particularly their list of questions, to see if their preparation is adequate and that their questions are appropriate and can sustain a discussion. Again, the purpose here is to guide the students and provide feedback. In other words, offer direction-- suggestions and/or corrections--only when asked or when it appears necessary to the students' success.

Class #2: Group leaders hand out copies of their articles and Discussion Outlines to their group members for them to read as homework. The purpose of this is to provide the students with a chance to build more schema so that they can speak about the topic more easily and discuss it in more depth. When I first started employing the RDG activity, I simply had group leaders hand out their articles and Discussion Outlines the day of the RDG. However, discussions were often stilted and did not flow well. I have found that since I started assigning the outlines as homework, my students discuss the topic much longer and in more depth, since they have read about it--and hopefully thought about it--beforehand.

Class #3: The RDG. As students are getting into their groups, hand out the Participant's Discussion Worksheet, Participant's Self-Evaluation and Discussion Leader's Self-Evaluation sheets to the students (see Appendix B6-B8). The Discussion Worksheet is used by all the students to record some of the answers their group gives to the leader's questions. This helps students of differing learning styles, especially those who are visual learners, to retain more of what they hear in the discussion. The worksheet can be filled out during the discussion or afterward, but is due the following class, regardless. The self-evaluation sheet is to be filled out as homework for the next class, as well.

---

[2] My main sources of articles for intermediate and advanced students are the English-language newspapers in Japan: *Japan Times, Daily Yomiuri, Asahi Evening News* and *Mainichi Daily News*. Each of these papers also publishes a weekly supplement containing bilingual articles of the week's important news stories, which are more accessible to lower-level students who otherwise would find news articles too difficult. All these publications are widely available at Kobe and Osaka kiosks and bookstores. Other possibilities include the *Hiragana Times, Nihongo Journal* and *Kansai Time Out*. These publications are available at Kansai-area English bookstores and through subscription. More advanced students can also handle articles from weekly news magazines, such as *Newsweek, Time* and *U.S. News & World Report*. These publications are widely available at university libraries, Kansai-area English bookstores and through subscription.

During the discussion, I walk around the class to monitor the discussions. This is basically to answer any questions that students may have and to ensure that all students are speaking only in English. I usually sit in on all of the groups for a few minutes as well, although I rarely speak, unless spoken to; I prefer to keep the focus of the discussion on the group leaders. My English-only rule is fairly strict, but I do allow my students to take a "*Nihongo* time-out" whenever they are stuck for something to say. In other words, if students find they cannot express themselves in English, they can call a "time-out"--just like in basketball--to ask for help from their group (or me) in translating something from Japanese to English. The reason why I feel this system is important is that it forces students to use as much English as possible, but still permits them to seek help in Japanese. After all, the goal of the language learning environment is to optimize the use of the target language; I think this system succeeds in doing just that.

The whole RDG activity typically lasts 35-45 minutes, although it can last as short as 20 minutes or as long as 90, depending on your students' levels and your class time. At the end of the RDG, I ask the group leaders to hand in their articles and Discussion Outlines so that I can write some comments on them. (I also use them to aid me in checking the responses students write on the Participants' Worksheets; otherwise, I cannot tell to what questions they are responding.)

If you have time, as a follow-up activity, you also might appoint one student in each group--not the leader--to present a short summary (2-3 minutes) of their group's discussion after the RDG ends. The value of this activity is that it gives students more speaking practice and a chance to practice their summarizing skills; it acts as a sort of informal speech. (Students who might have said they can not give a speech in English will be stunned when you tell them they just did!)

Class #4: Students hand in their Participants' Worksheets and Self-Evaluations. The Self-Evaluations ask students to assess themselves on 20 criteria in four different areas: Preparation, Initiative (Leadership), English Usage and Facilitating & Moderating (for the leaders) or Participation & Cooperation (for the participants). This form was constructed based on a "Participation Wheel" idea developed by my colleague Maricel Santos--which she adapted from elementary education literacy materials (Henning-Stout, 1994, p. 112-119)--and materials developed by Schmidt and Wong (1993, p. 52-63) for San Francisco State University's American Language Institute (see below). The reason for these four different assessment areas is that it helps to give the students a more holistic picture of their overall performance, rather than focusing on what they did during the discussion only. The discreet points simply provide the students with more specific criteria and objectives, so that they know what they need to do in detail. As it is fairly complex, though, you will probably want to opt for a simpler assessment form for lower-level students; but it is still possible to use the same system, I think.

Again, I take the students' self-assessments at face value; as with the oral quizzes, the criteria for assessment are quite specific, so there is little chance the students will not understand them correctly, particularly if you have explained the content and modeled it for them previously.

Comments

Although this is an activity which has a clear set of objectives, when I first started using the RDG three years ago, I soon found that the way it was originally structured had several weaknesses. First, the procedure was quite complex--at least to the students, most of whom had never done an activity like it and consequently had a hard time following all the guidelines. Much of this problem was solved when I ran across a set of preparation guidelines in a coursebook published by San Francisco State University's American Language Institute, with materials contributed by Schmidt and

Wong (1993, p. 52-63), for an activity very similar to the one I had developed. Once I had adapted and refined some of the ALI materials for my class, my students were much more consistent in following the guidelines and preparing well. However, I found that I still needed to model the RDG for my classes before letting the students take charge, even with advanced students and returnees.

Second, at first the RDG did not offer my students a good sense of closure; there was no attempt made at having students do any self-reflection or self-evaluation after completing it. Their performances were at first evaluated only by me. But I found it very hard to assess each student accurately, since there were 3-5 groups working simultaneously in a class of 20-30 students. This prompted me to develop a self-assessment instrument that I could incorporate into the end of the activity. I have found that self-assessment provides students with a better sense of completion--especially of having worked toward and reached a goal, rather than simply having had a discussion.

Having now worked out these kinks, the RDG is now an integral--and very valuable--part of most of my courses, at every level. In fact, many of my students have written in their course evaluations that the RDG was the most helpful activity for them during the entire course. One student wrote, for example: "The Discussion group was so helpful, because it helped me to improve my speaking skill." All of my students have enjoyed the extensive speaking practice, and many have said that they now feel more comfortable and confident speaking in English. Again, I think this is due in part to the fact that they get to speak with their peers, in the protection of a small group, so there is less "performance anxiety" when they speak. Also, students have some knowledge of the subject beforehand, so they have a chance to recycle much of the new language they have learned.

As for the self-assessment technique, many students have told me that although it was hard for them to get used to at first, they preferred it because they were forced to think about what they had done and learned, and they also knew what criteria they had to meet. I feel that as an assessment tool the self-assessment strategy is quite valid because it is "so closely linked to the phenomena under observation that it is 'obviously' providing valid data" (Kirk and Miller, 1986, as cited in Huerta-Macías, 1995, p. 9).

In addition, as with the oral quizzes, the students' self-assessment scores were not inflated--in fact, anything but: most students gave themselves scores in the B and C range (60 to 80%); several even failed themselves, just as I would have failed them. There were some students, in fact, whom I thought were unduly self-critical and harsh with their self-assessments. My one regret is that I did not have time to sit down with each of them and discuss their rationale for their scores. However, as John Cowan (1988) writes, it is important to remain a detached observer and not make judgments on the students' marks (p. 202).

There are various advantages to the RDG activity, as well. First, as with oral quizzes, it is beneficial to all your students, even if there is a large disparity of skill levels among the students in your class. Students at all levels of proficiency can still get something from it. Second, the RDG can be utilized in many types of classes: reading, writing, listening, speaking or integrated skills, since the RDG contains components of all these skills. Third, it works well with large classes, even those which have diverse proficiency levels (as explained above). Fourth, it is learner-centered. Fifth, the RDG fits in well with a theme-based curriculum, since it can be related to themes which are already being worked on in class. And lastly, the RDG provides a lot of speaking practice; so, the RDG is another good way for students to get more practice in a way which is conducive to language acquisition.

3. Debates

A third common classroom activity in which you can easily employ both peer- and self-assessment is debate. There are many different formats used for both formal and informal debating, so I will not review them now. Instead I would like to briefly review the format I have used in my classes and focus on the assessment procedures.

Background

As with the RDGs, I have used formal debates in several of my intermediate and advanced classes which have 20-30 students. The most effective way that I have found to conduct a number of team debates in a large class without producing a lot of idle, non-task time--and student boredom--is to use peer-assessment and self-assessment strategies for evaluating the debates. I also divide students into teams of three or four instead of two, so that there are fewer actual debates; this helps decrease the preparation work load on each team and provides a little more 'safety in numbers,' which Japanese students tend to find more comfortable.

As with the RDGs, I still have not solved the problem of how to divide the students into teams--by proficiency or randomly. However, since debate is a team activity, I think it is important that all the teams contain a similar "amount" (if you can call it such) of proficiency. In other words, I try to divide the students into teams of mixed proficiencies that somehow equal out--for example, team A might have one high-level, one middle-level and one lower-level student, while team B might have three middle-level students--so that each *pair* of teams is closely matched in overall proficiency. This system has produced the greatest amount of parity, so that the differences in the debate performances can be measured more in terms of preparation and proper use of debating strategies.

The Setup

As with the RDGs, before my students engage in a debate, it is important that they build some schema--background knowledge and vocabulary--for the chosen topic. I follow a similar thematic scheme as with the RDGs, including a film clip, a reading and/or an informal discussion to activate students' schema and build more topic knowledge and vocabulary. I have also found it helpful to have students watch a videotape of a successful student debate from another class as a model--although this might be hard if this is the first time you are using debate in your class. (Check with colleagues!) Teachers should also supply students with handouts on proper debate etiquette and tips on how to be persuasive.

Choosing an appropriate topic is also tricky. I have discovered that when pushed, students are quite capable of debating on a topic that they (or you!) thought was too hard for them. Recently, for example, my intermediate students have debated on public smoking, euthanasia (mercy killing) and the consumption tax. Obviously, lower-level students may have to debate on topics more accessible to them, but do not underestimate their abilities!

Students also need lots of time to adequately prepare for a debate, particularly if you are going to require them to do any outside reading or research. I typically give my lower or intermediate students two articles (or more) about a topic--usually ones with contrasting viewpoints--to use as a reference guide for constructing their debate strategy. In my advanced classes, students do the research themselves.

The Procedure

Step 1: Background: How much time students require depends on how in depth you want to go with the activity, and how formal you want the debates to be. If, as I

mentioned above, you want to include instruction on proper debate etiquette and strategy, this itself might take one full class. Going over the debate format will probably also take half an hour--maybe more.

Step 2: Preparation: I usually assign the reading of the articles on the topic as homework. However, students will still need at least one class period--and preferably two--to prepare their strategies, including planning questions to ask, ways to attack the opposing team's arguments and ways to defend their own team's position (for which the students should have handouts, as well). During this time, I circulate around the room and help students with the wording of questions or statements and strategy. I also tend to allow students to speak a bit more *Nihongo* than normal, but I still require them to call a "time-out". Generally, teachers simply have to make certain that students are on task, working together equally, and that they are planning an effective strategy. I think that with debate, teachers can--and should--provide more explicit direction than with RDGs or oral quizzes.

Step 3: The debates: Students are grouped into their teams, and usually I pick the teams by lot the class before, so students know when their team is scheduled to go. The debate teams sit at tables in front of the class which are placed diagonally, facing each other, with the moderator's table between them (facing the audience, though) so that the participants can see each other and the audience can see both teams. The student audience hence becomes the judges for the debates. As the teams move to their respective tables, I distribute the Debate Peer Evaluation and Debate Self-Evaluation forms (see Appendix C2-C3).

Usually, I act as debate moderator, but there is no reason that students could not do it. The only requirement is that the moderator be very strict with keeping time--although with an inexperienced or low level class, you may want to be more flexible with the time restraints. With three students per team, using the format listed on Appendix C1, each debate takes approximately 30 minutes, which includes time for students to fill out evaluations afterwards. Therefore, in one class period, 18 students can take part in debates. So, most classes can conduct their debates in one or two class periods. In a large class of 40-50 students, it might be desirable to put students into teams of four, so that you do not spend more than two class periods on the debates.

Step 4: Evaluation: After each debate, all students are required to complete an evaluation: the student-judges complete the peer-assessment sheet, while the participants complete the self-assessment form. There are six specific statements that the judges must evaluate, followed by an overall evaluation. Debate participants must evaluate themselves on seven specific criteria relating to both their preparation and performance, using a four-point graded scale (1 = weak; 4 = strong). The eighth statement is their own evaluation of which team won. On the bottom of the page, I also ask both judges and participants to write short answers about the strongest and weakest points of each team's performance. This hopefully leads the judges to think more deeply about the debate topic and strategy, and the participants to reflect more on their performance. It also gives all students more practice writing their own opinions quickly and concisely. Finally, I collect all the evaluations and tabulate the votes, with the winning team members getting bonus points.

Comments

I have used peer- and self-assessments for evaluating debates only one year, so I am still learning much about the process and the effectiveness of the procedure. There are a few conclusions I have reached, however. First, the advantages of using debate are similar to those listed above for RDGs.

Second, of all the activities I use in my English classes, debate seems to be the most difficult and troublesome for my students. This is due largely to the lack of familiarity Japanese students have with this type of activity--and this Western style of learning. Great patience must be exercised in preparing for the debate. Do not expect your students to stage a smooth, well-paced and rapid-fire debate, either. Even my most advanced students have struggled with debating. However, all of my students--particularly those with overseas experience--have said that the debate is a valuable learning activity, as it provides them with a chance to think more about current topics. One student wrote:

> *The debate was a really interesting activity though it was hard for me. When I went on [a summer] exchange program to England last year, there were many students from different countries. We had discussion very often about many topics. I was ashamed, not because of my poor English ability but because I did not have any opinions about those topics. I had not thought of them before. So, your debate made me to think [sic] more about international issues and Japanese culture.*

Third, the debate offers students a chance to express their ideas and opinions, but forces them to do it in a limited time frame. This again is something to which Japanese students are not accustomed. The peer and self assessments, therefore, prompt students to reflect on their own and their peers' abilities to express their thoughts and opinions quickly and effectively.

Fourth, as an assessment tool, the Debate Self-evaluation is easy for students to understand and complete, although its value as a measurement of language learning progress is unclear. It is useful, however, as a means of self-reflection. As with both the previous activities, students' self-assessment scores were very similar to my own assessments of their performance, indicating that the comparative simplicity of the assessment criteria did not tempt (or at least cause) students to inflate their scores.

Finally, in the future, it might be more helpful to ask students to complete a detailed assessment about the debate, such as what new language they learned, what they learned about the topic, and more in-depth questions about what they found out about their own speaking skill.

4. Overall Semester Class Participation Self-Assessment

Finally, I would like to briefly touch on a semester self-evaluation that I have just started using this year. This idea, therefore, is still fairly germinal, but I believe that it can still be discussed in some detail.

Background

In the Intensive English Program, class participation in the Intermediate and Pre-Advanced classes accounts for 30% of the student's grade (this is a standard). In 1995, Language Center instructors--including myself--got together to try and standardize the way this participation score would be determined. The result is the grid that appears on Appendix D1. We concluded that there were six different areas (or "categories") of participation that we wanted to measure: preparation, initiative (self-motivation), English use, cooperation, attentiveness and overall contribution to class--each with equal weight. We felt that this distribution of values presented us with a more holistic picture of the student's overall participation. Until this point, in fact, we did not even have a guide for evaluating class participation, so the grid at least presented us with a more systematic--and fair--way to evaluate students holistically than the method we had been using.

From this grid, Maricel Santos developed a more sophisticated version of assessing participation in 1996, the Participation Wheel, which she adapted from elementary education literacy materials (Henning-Stout, 1994, p. 112-119). The Participation Wheel contains the same four areas of assessment as mentioned in the RDG section, above (also see Appendix D3). I then decided that if my students could assess themselves on a smaller scale, as with the above activities, there was no reason they should not be able to assess themselves on a larger scale, particularly when the assessment instrument being used was one with which they were already familiar. Adapting the Participation Wheel to the level of my students (low- to high-intermediate) required only some fine tuning, as well. I made just a few minor changes to Santos' original wheel in order to end up with 30 discreet points on which students could assess themselves.

The Procedure

Quite simply, you can either have the students complete the self-assessment as their final homework for the final class of the semester, or have them complete it on the final day of class. It takes approximately 20 minutes to complete. The advantage of having them do it as homework is that the students can put more thought into their scores.

Again, take the students' scores at face value, and make sure that they understand that whatever score they come up with will be their mark, no questions asked.

Comments

Initially, my students were shocked (dumbfounded would be more accurate) that I was giving them the sole responsibility for evaluating their participation in the class--especially since it accounted for 30% of their grade. As I explained to them, however--to paraphrase Murphey (1995)--the most important evaluation is ultimately what they think of themselves.

To my equal shock and amazement, all 45 of the students who used the Participation Wheel to assess themselves in three of my classes evaluated themselves at or *below* my own private assessment of their class participation. Perhaps I had some unusually honest students in my classes, but I prefer to think that they respected the responsibility that had been bestowed upon them and responded by honoring my trust in them. As with the RDGs, most students gave themselves scores in the B and C range (60 to 80%), a few gave themselves high marks, and several failed themselves. Again, there were some students whom I thought were unduly self-critical and harsh with their self-assessment, and I wish I could have discussed their rationale for their scores.

There is one weakness with the Participation Wheel, though, in its current form: the original participation grid we used is on a graded scale, but the wheel is strictly linear--yes or no. Upon reflection, I feel that students should be able to choose points along a graded scale for each discreet point on the wheel, just as we as teachers could do with the grid. This may end up not making much difference in the overall score, though--perhaps a point or two. But, when participation accounts for 30% of the grade, one or two points can make a difference in the grade a student receives. In the future, therefore, I will offer students in one of my classes a six-point scale version of the Participation Wheel, to see if (or how much) difference a graded scale makes.

Regardless of this weakness, though, this first initial trial with participation self-assessment has shown me that when given the responsibility, even on a large scale, students will take it, and can be relied upon to accurately and honestly assess their own performance.

CONCLUSION

Self-assessment is still a relatively novel teaching technique to me (and most other teachers), and I am still learning how it works and how and when to use it. As mentioned earlier, it is by no means a panacea for English education, but it does offer teachers more assessment options, and it does get students more involved in the language learning process. As stated earlier, I believe students benefit more from the self-awareness of the learning process and the self-development they procure by taking responsibility for their own learning and the assessment of it.

The hardest aspect of self-assessment for me, as a teacher, has been letting go of the control of the assessment because I am, in essence, giving up part of my control over the students. That is a big adjustment, but one which can be made. As Cowan (1988) states, it is essential to the success of self-assessment that students regard the teacher's comments as no more than the reaction of a detached outside observer, to whom they are free to respond or ignore, as they wish (p. 202).

Perhaps my feelings about using self-assessment can best be summed up by Cowan (1988), who concludes:

I believe the most important point I have learned about facilitating self-assessment is that I must never allow myself to formulate my own judgment of the mark or rating which a student should receive... [W]ithout access to all that the learners know of their learning, any judgments of mine can only be shallow and unhelpful. (p. 202)

REFERENCES

Cowan, J. (1988). Struggling with self-assessment. In D. Boud (Ed.), *Developing student autonomy in learning* (pp. 192-210). London: Kogan Page.

Gardner, D. (1996). Self-assessment for self-access learners. *TESOL Journal, 5,* (3), 18-23.

Hakner, S., & Cutolo, A. (1997). Reflections on an assessment dilemma. *TESOL Matters, 6,* (6), 10.

Henning-Stout, M. (1994). *A new way of thinking about learning.* San Francisco: Jossey-Bass Publishers.

Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5,* (1), 8-11.

McNamara, M.J., & Deane, D. (1995). Self-assessment activities: Toward autonomy in language learning. *TESOL Journal, 5,* (1), 17-21.

Murphey, T. (1995). Tests: Learning through negotiated interaction. *TESOL Journal, 4,* (2), 12-16.

Schmidt, J., & Wong, J. (1993). Seminars. In *Academic listening and speaking skills: Course materials for Level 48/OCS Listening.* San Francisco: San Francisco State University.

VOCAB LIST for ORAL QUIZ # ____ Name_____

EXAMPLE

qualm (or qualms)	TYPE: **noun**	(Source) **JOHN**
Meaning:	**concern, worry; especially an uneasy feeling about whether doing something is right.**	
Example:	**John has no qualms about giving his IEP students lots of homework. In fact, he is quite comfortable with it.**	

Entry # 1

	TYPE:	(Source)
Meaning:		
Example:		

Entry # 2

	TYPE:	(Source)
Meaning:		
Example:		

Entry # 3

	TYPE:	(Source)
Meaning:		
Example:		

Entry # 4

	TYPE:	(Source)
Meaning:		
Example:		

(NOTE: This sample contains only 4 entries; normally students get 5 sheets, totalling 20.)

ORAL VOCABULARY QUIZ INSTRUCTIONS

INSTRUCTIONS: When you reach a comfortable place where you want to take the quiz, give your "Vocabulary Study List" sheet to your partner. Your partner will then ask you about **all 20** entries you have on this list. When your partner has finished quizzing you, then you will change roles, and you will quiz your partner.

PART 1: WHEN YOU ARE ASKING QUESTIONS

Ask your partner one entry at a time, AT RANDOM! Do not simply go down the list! Your partner should start to answer in **less than 5 seconds**; if they do NOT start to answer in **5 seconds**, then give them a "BUZZER" sound—this means that their time is up! But, when your partner starts to speak, give them as much time as they need to answer the question completely.

This is how you should ask the questions:
"Entry number __ : please tell me what _____ means and use it in a sentence."

After listening to your partner's answer, talk to them and discuss the number of points they deserve for the response, and enter this on the mark sheet. When your partner has finished explaining all 20 entries, add up the total number of points they earned, including penalty points for speaking *Nihongo*, and write this score on the bottom line of the mark sheet. Then give your partner's Vocab Study List back to them.

PART 2: WHEN YOU ARE ANSWERING QUESTIONS

You must give the definition and an example sentence for each entry, IN ENGLISH ONLY, and in COMPLETE sentences! (No one- or two-word answers!)

☞ Remember that **you must start to answer in less than 5 seconds**! If you are having trouble remembering the vocabulary, try to "buy time"—just say ANYTHING in English, like, "Hmm—that's an interesting question. Many students probably don't know that one. I studied very hard myself, but I'm having trouble remembering it..." and so on. Even if you cannot remember the word, you can **still** earn one point for it just for speaking in English quickly, and for 4 or 5 sentences without pausing!

(4) You have **25 minutes only** to complete the quiz! Please do not talk to any other KGU students you may meet while taking the quiz! You have a time limit, and I will penalize you each **5 points per minute** for every minute you go over the limit!

(5) After you have finished quizzing each other, give me your Vocab Study List and your mark sheets. I will check your list to make sure you have the correct meanings and a clear example.

☞ **REMEMBER! On your list, each entry MUST have a CLEAR definition and a COMPLETE EXAMPLE SENTENCE, and you MUST list the TYPE of word it is and which unit or activity it is from!**

ORAL VOCABULARY QUIZ MARK SHEET

Your Name: _____ Your Partner: _____

SCORING GUIDE for each QUIZ: Ask your partner about **all 20** of the entries on their list! Give your partner a score for each word, and then a total score, based on this system:

- Clear and correct explanation of the entry: (including **definition** and an **example sentence**) — 4 points □
- Speaking quickly & continuously in English: — 1 point °

TOTAL for each entry = 5 points
x 20 words
= **100** points

- PENALTY POINTS for speaking JAPANESE: — 1 point per word!*

= GRAND TOTAL

□ = If the answer is MOSTLY correct, give your partner 3 points
If there is some information missing, give your partner 2 points.
If the answer has several mistakes, give your partner 1 point.
If the answer is WRONG or is not given in time, give your partner 0 points.

° = If your partner has any long pauses in their answer, do NOT give them this point!

* = Deduct 1 point from your partner's score for each time they speak ANY *Nihongo!*
—This includes even small words, like "ano", "ma", "ne", etc.!

YOUR QUIZ SCORE (filled in by your partner!!!)

ENTRY #	POINTS	ENTRY #	POINTS
1	(5)	11	(5)
2	(5)	12	(5)
3	(5)	13	(5)
4	(5)	14	(5)
5	(5)	15	(5)
6	(5)	16	(5)
7	(5)	17	(5)
8	(5)	18	(5)
9	(5)	19	(5)
10	(5)	20	(5)
Penalty	(-1 point per word!)		

TOTAL = _____ (out of 100) Your signature _____

Your partner's signature _____

INTRODUCTION

What *is* a "Reading Discussion Group"?

Three skills I want you to develop in this course are: ① expressing your ideas & opinions clearly in English; ② leading a discussion; and ③ critical thinking skills. Group discussions are an excellent way to develop these skills. So, this semester, you will prepare for and participate in a series of organized group discussions, called Reading Discussion Groups (RDGs), which are also known as *seminar discussions* or *seminars*.

A *seminar* is a kind of class in which the students do most of the talking, not the teacher. The student-led seminar is a common format for many college and graduate school courses in Western countries—especially the U.S. Many educators value it because it encourages students to think critically and to develop a deeper understanding of the subjects they study. These skills are also essential for the Advanced courses.

More specifically, in a *seminar discussion*, students research information on a topic, report on their findings, and lead an organized class discussion on the topic. The topics for the RDGs in Pre-Advanced will be chosen by you, but I will supply you with the materials you need to organize & lead a seminar discussion: newspaper & magazine articles, or video clips. So, *you* will be expected to summarize the article, study the vocabulary & be able to explain it to your group, and think of discussion questions about the topic so tou can lead a discussion about it.

The most important aspect of the RDG is **participation**! Students are evaluated by both the quantity and quality of their participation in these group discussions. They must carefully analyze & evaluate the information they receive, and are expected to reflect & comment on it. So, just following the discussion topic is not good participation. Students must actively exchange ideas and opinions. So, participating in a RDG involves both a good presentation & interactive discussion. In addition, since I believe that you are your own best critic, evaluations for the RDGs will be done both by me and by yourself. On the following pages is a description of the format for the RDGs.

WHAT KIND OF LEADER ARE YOU?

Materials adapted from: Henning-Stout, M. (1994). *A new way of thinking about learning*, 112-119. San Francisco: Jossey-Bass Publishers.

Reading Discussion Group Format

I. GIVE A SUMMARY OF THE ARTICLE (about 5-7 minutes)

A. Background / Introduction

1. State the title of your article & tell where the article comes from (name of publication & writer)
2. Explain your RDG topic.
3. Tell the group why the topic is controversial or interesting. Provide some general background information about your topic, if necessary.

B. Review the article itself

1. Identify the 3-4 **key ideas** from your article which best explain or clarify your topic.
2. Tell the class about your first key idea.
   - *Describe your first key idea by providing specific examples from the article.*
   - *Avoid reading directly from the article.*
   - *Paraphrase your examples.*
3. Tell the class about your next key idea.
   (Same as above.)
4. Go over the list of special vocabulary on page 4. Answer any questions about other vocabulary that group members may have.

C. Transition to Discussion

1. Restate why the article is interesting or important to the topic
2. Transition to discussion segment

II. DISCUSSION OF TOPIC (20 minutes)

A. Review the **Information Questions*** (2 or 3). These are questions based on information in the article.
B. Go over the **"Thought" Questions*** (2 or 3). These are questions which ask for the group members' ideas & opinions about the topic.
C. Conclude the discussion. Review some of the main points that the group brought up.

***The leader should also create a list of 2-3 <u>additional</u> questions for the discussion, in case your group finishes discussing these questions early!**

Materials adapted from: Herrmann-Stout, M. (1994). *A new way of thinking about learning.* 112-119. San Francisco: Jossey-Bass Publishers.

Discussion Outline

LEADER: _____

Directions: **LEADER**, complete this outline; then copy it off and give it to your group members.

I. SUMMARY OF ARTICLE (Estimate: _____ minutes)

A. Introduction

1. Article title, source (publication) & author (if known):

2. Explanation of topic (background):

3. Importance / interest of topic:

B. Review of the article

1. Key point: _____

2. Key point: _____

3. Key point: _____

Materials adapted from: Schmidt, J., & Wong, J. (1993). Seminars. In *Academic listening and speaking skills: Course materials for Level 48/OCS Listening*, 52-63. San Francisco: San Francisco State University.

C. Review of Important Vocabulary (list only 5 - 10)

- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =
- _____ ( ) =

D. Transition to Discussion

1. Restatement of topic's importance

2. Transition sentence (example: *"Now I would like to move on to the discussion questions."*)

Materials adapted from: Schmidt, J., & Wong, J. (1993). Seminars. In *Academic listening and speaking skills: Course materials for Level 48/OCS Listening*, 52-63. San Francisco: San Francisco State University.

Leader: _____ Group: _____ RDG # _____

II. DISCUSSION OF TOPIC

Information Questions

1. Question: _____

_____

2. Question: _____

_____

(3). Question: _____

_____

"Thought" Questions

1. Question: _____

_____

2. Question: _____

_____

(3). Question: _____

_____

Materials adapted from: Schmidt, J., & Wong, J. (1993). Seminars. In *Academic listening and speaking skills: Course materials for Level 48/OCS Listening*, 52-63. San Francisco: San Francisco State University.

Participant's Discussion Worksheet

Directions: Fill out the answers to these questions during your group discussion.

Your Name: _____ Leader's Name: _____

☞ What are some answers your group gave to the leader's discussion questions?

Information Questions

_____

_____

_____

_____

_____

_____

"Thought" Questions

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Materials adapted from: Schmidt, L. & Wong, J. (1993). Seminars. In *Academic listening and speaking skills: Course materials for Level 48/OCS Listening*, 52-63. San Francisco: San Francisco State University.

Participant's Self-Evaluation

Your Name: _____ Leader's Name: _____
RDG # _____ Your Group: _____ Date: _____

Directions: Think about your participation in the RDG. Rate yourself, based on the following criteria, by choosing the items which best represents how well you did. **PLEASE BE HONEST!**

Preparation
0. Didn't read the materials before the RDG
◊ Read the article before the RDG class
◊ Looked through the new vocabulary & discussion questions before the RDG
◊ Thought about various possible answers to the discussion questions
◊ Thought about how to explain my ideas & opinions about the topic in English
◊ Tried to relate the topic we discussed to my own experiences and knowledge

Participation & Cooperation
0. Didn't pay attention or participate much
◊ Paid attention to the discussion at all times
◊ Asked questions or made comments to keep the discussion going
◊ Responded to comments made by others
◊ Tried to think about & analyze the ideas & opinions people presented
◊ Provided positive and constructive feedback to others

Initiative (Leadership)
0. Didn't speak at all during the discussion
◊ Spoke several times during the discussion
◊ Initiated some of the discussion
◊ Offered my own opinion
◊ Asked for help when I couldn't understand something during the discussion
◊ Offered my help when others couldn't understand something during the discussion

English Usage
0 Spoke often inJapanese
◊ Spoke in English almost all of the time
◊ Encouraged others to use English
◊ Made an effort to think in English
◊ Tried to express my idea, even if I couldn't remember the exact English word or phrase
◊ Wrote down new words, phrases or idioms that I heard during the discussion

Directions: Give yourself a participation score based on the scale above.

18-20 = Excellent
16-17 = Good
14-15 = Not Bad/So-so
12-13 = Barely Acceptable
11 - 0 = Unacceptable/Poor

Self-Evaluation Score on Your Participation:

_____

Materials adapted from: Henning-Stout, M. (1994). *A new way of thinking about learning*, 112-119. San Francisco: Jossey-Bass Publishers.

Discussion Leader's Self-Evaluation

Your Name: _____

Your Group: _____ RDG # _____ Date: _____

Directions: Think about your participation in the RDG. Rate yourself, based on the following criteria, by choosing the items which best represents how well you did. **PLEASE BE HONEST!**

Preparation

0. Didn't prepare the materials before the RDG
◊ Prepared a summary of article's key ideas
◊ Made a list of new vocabulary & discussion questions before the RDG
◊ Thought about various possible answers to my discussion questions
◊ Thought about how to explain my ideas & opinions about the topic in English
◊ Tried to relate the topic we discussed to my own experiences and knowledge

Facilitating & Moderating

0. Wasn't able to keep the discussion going.
◊ Made sure group members understood the article, vocab & my discussion questions
◊ Asked questions or made comments to keep the discussion going
◊ Tried to think about & analyze the ideas & opinions group members presented
◊ Provided positive and constructive feedback to others
◊ Was able to successfully clarify, rephrase & summarize information from the discussion

Initiative (Leadership)

0. Didn't lead the discussion at all
◊ Initiated some of the discussion
◊ Kept the discussion focused on the topic
◊ Offered my own opinion about the topic
◊ Made sure everyone in the group participated
◊ Offered my help when group members couldn't understand something

English Usage

0. Spoke often inJapanese
◊ Spoke in English almost always
◊ Encouraged group members to use English
◊ Made an effort to think in English
◊ Tried to express my idea, even if I couldn't remember the exact English word or phrase
◊ Wrote down new words, phrases or idioms that I heard during the discussion

Directions: Give yourself a participation score based on the scale above.

18-20 = Excellent
16-17 = Good
14-15 = Not Bad/So-so
12-13 = Barely Acceptable
11 - 0 = Unacceptable/Poor

Self-Evaluation Score on Your Leadership in this RDG:

_____

Materials adapted from: Henning-Stout, M. (1994). *A new way of thinking about learning*, 112-119. San Francisco: Jossey-Bass Publishers.

Discussion Leader's "To Do" Checklist

☑ **1. Pick up a copy of the article from your instructor.**

☐ **2. Read the article and prepare a summary of its main ideas**

Read through the entire article carefully. Make a list of the main ideas—there should be at least three! Check any vocabulary you don't know; if you think it is important to understanding the text, then make a note of these words and be sure to explain them to your group.

☐ **3. Write up a vocabulary list**

Write a list of 5-10 important vocabulary words *and their definitions* from the article and be able to explain these—and any other vocab words from the article—to your group members.

☐ **4. Write up a list of discussion questions**

Think of at least **4** questions for discussion. Two of the questions should be **INFORMATION** questions, which are based on information in the article. The other two (or more) should be **THOUGHT** questions, which ask other students to give their opinions or present their own ideas about the topic. *Think carefully about your questions and how your group may respond to them!* This list of questions should be written on the **Discussion Outline** form. Copy off this form for your group members.

☐ **5. Copy off all materials (article, vocab list, and discussion outline)**

You are responsible for making copies of the article & your outline for your group members! If you would like any advice or assistance from me, or would like to make copies in the Language Center (they're free), please come see me in the Language Center office!

☐ **6. Give all materials to your group members**

All materials MUST be copied off and distributed to your group members *the class before you lead the discussion!*

☐ **7. Lead the discussion about your topic; then, hand in *Discussion Outline***

☐ **8. Complete *Self-Evaluation* Form & hand in at the next class**

Materials adapted from: Henning-Stout, M. (1994). *A new way of thinking about learning*, 112-119. San Francisco: Jossey-Bass Publishers.

DEBATE FORMAT

I. Welcome	The moderator will begin the debate by announcing the topic of the debate and and introducing the teams and participants.
II. Introductory Statements	PRO team member #1 gives an introductory statement, stating the reasons for their team's position. (1.5 mins) CON team member #1 gives an introductory statement, stating the reasons for their team's position. (1.5 mins)
III. Exchange of Questions and Answers	A CON team member #2 asks a PRO team member a question. PRO team member confers with teammates. (30 seconds) PRO team member answers. (1.5 minutes) CON team member #2 confers with teammates. (30 seconds) CON team member #2 rebuts. (1 minute) B PRO team member #2 asks a CON team member a question. CON team member confers with teammates. (30 seconds) CON team member answers. (1.5 minutes) PRO team member #2 confers with teammates. (30 seconds) PRO team member #2 rebuts. (1 minute) C CON team member #3 asks a PRO team member a question. PRO team member confers with teammates. (30 seconds) PRO team member answers. (1.5 minutes) CON team member #3 confers with teammates. (30 seconds) CON team member #3 rebuts. (1 minute) D PRO team member #3 asks a CON team member a question. CON team member confers with teammates. (30 seconds) CON team member answers. (1.5 minutes) PRO team member #3 confers with teammates. (30 seconds) PRO team member #3 rebuts. (1 minute)
IV. Closing Statements	PRO team member #1 gives their closing statement. (2 minutes) CON team member #1 gives their closing statement. (2 minutes)

DEBATE PEER EVALUATION FORM

Directions: Listen to the debate carefully and fill out this form. Put a check or 'X' in the box for the team which did the best job for each category.

	PRO	CON	NEITHER
This team's introductory statement more clearly stated their position, and was more convincing.			
This team's questions were more clearly stated and challenged the opposing team's position.			
This team's answers to the opposing team's questions were clearer, more logical and more persuasive.			
This team's concluding statement more clearly stated their position, and was more convincing.			
This team seemed to be more organized and more cooperative with each other.			
This team was more able to persuade me to accept their position.			
I THINK THE WINNER OF THIS DEBATE WAS:			

- What was the strongest point about the PRO team's debate? Why?

- What was the strongest point about the CON team's debate? Why?

- What was the weakest point about the PRO team's debate? Why?

- What was the weakest point about the CON team's debate? Why?

DEBATE SELF-EVALUATION FORM

Your Name _____ Your Team: PRO CON

Directions: Please circle the number which best represents how you think you and your team performed in the debate.

	weaker <----------> stronger			
1. I was able to present my team's position clearly.	1	2	3	4
2. I was able to express my ideas in English quickly and clearly.	1	2	3	4
3. I was able to answer the opposing team's questions clearly and logically.	1	2	3	4
4. My team was able to present its position clearly.	1	2	3	4
5. My team's introductory and concluding statements were clear and convincing.	1	2	3	4
6. My team's questions were clear and challenged the opposing team's position.	1	2	3	4
7. My team was well organized and cooperated well with each other.	1	2	3	4
8. My team won this debate.	NO 0	It was a tie 1		YES 2

TOTAL POINTS: _____

• What was the strongest point about your team's presentation? The other team's? Why?

• What was the weakest point about your team's presentation? The other team's? Why?

CLASS PARTICIPATION GRID (1995)

CATEGORIES:	SCORING: (by category)
• Preparation for class	5 points = **almost always**
• Leadership/Self-Motivation	4 points = **usually**
• English Use	3 points = **often**
• Cooperation with Others	2 points = **sometimes**
• Active Listening	1 point = **rarely**
• Contribution to Class (Discussions, etc.)	0 points = **never**

BEST COPY AVAILABLE

CLASS PARTICIPATION SELF-EVALUATION

Your Name _____ **Student ID#:** _____

DIRECTIONS: Please mark the choices below which you feel accurately reflect your participation in this course this semester:

BASICS

◊ 1. I showed an interest in improving my English this semester.
◊ 2. I was able to make a list of personal goals for this class.
◊ 3. I was able to meet some or all of these goals in this class.

PREPARATION for CLASS

◊ 1. I brought my text & other materials to class (almost) every day.
◊ 2a. I did my homework regularly & took notes in class to remember new vocab
◊ 2b. I prepared for class projects & activities.
◊ 3a. I reviewed class material and was able to explain interesting ideas or meaning.
◊ 3b. I tried to relate what we did in class to my own knowledge & experiences.

INITIATIVE & LEADERSHIP

◊ 1. I willingly took part in classroom activities.
◊ 2a. I tried to speak in class several times each day.
◊ 2b. I tried to begin discussion with my classmates.
◊ 3a. I offered my own opinion.
◊ 3b. I responded to comments made by my classmates and teacher.
◊ 4a. I asked for help when I didn't understand something we were studying.
◊ 4b. I was able to work on my own (to prepare for projects, etc.), on materials I selected.
◊ 5. I was able to challenge myself to choose challenging (difficult) activities and projects.

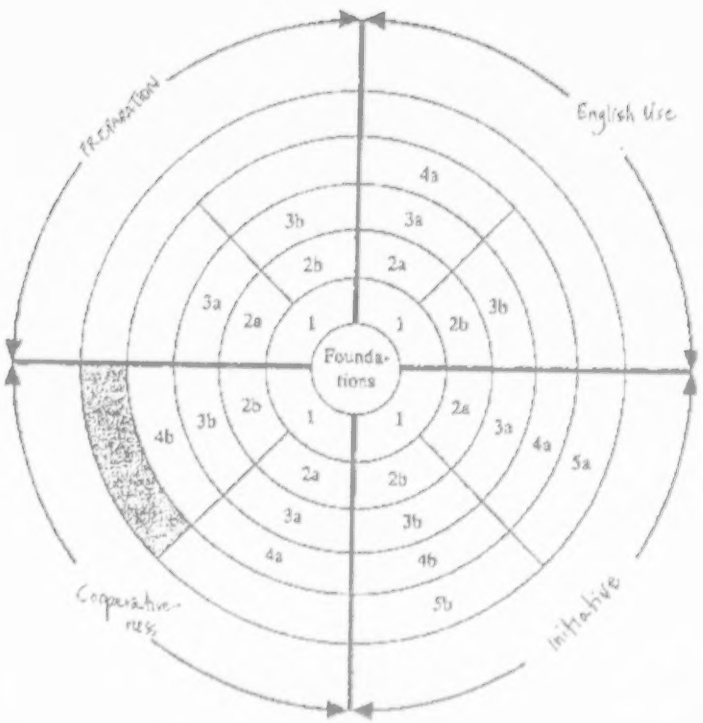
USING ENGLISH IN CLASS

◊ 1a. I spoke in English most of the time in class.
◊ 1b. I spoke in English (almost) all the time in class.
2a. I gradually felt more comfortable speaking in English in front of others.
◊ 2b. I tried to get my message across, even if I couldn't remember the exact words in English.
◊ 3a. I encouraged my classmates to use English.
◊ 3b. I made an effort to think in English as much as possible.
◊ 4. I wrote down new words, phrases or grammar points that I heard in class.

COOPERATION

- ◊ 1. I paid attention in class as much as possible.
- ◊ 2a. I tried to ask questions connected to what we were studying or talking about.
- ◊ 2b. I was able to work well with my classmates in a group.
- ◊ 3a. I was able to stay on task (not goof off).
- ◊ 3b. I tried to have different partners for different class activities.
- ◊ 4a. I shared talk time (I didn't dominate the conversation).
- ◊ 4b. I tried to provide positive and helpful feedback to my classmates.

TOTAL POINTS _____



Materials adapted from: Henning-Stout, M. (1994). *A new way of thinking about learning*, 112-119. San Francisco: Jossey-Bass Publishers.

Chapter 6

Interactive Vocabulary Quizzes: An Alternative Form of Assessment Based on Multiple Intelligence Theory

GINA KEEFER

MULTIPLE INTELLIGENCE THEORY

Our roles as educators are changing as current research continuously evolves in the study of intelligence. Howard Gardner's increasingly-recognized work has expanded our thinking from a simplistic to a much more holistic and complex understanding of intelligence which recognizes the unique spectrum of talent found within each individual (Gardner, 1993). Gardner's theory of multiple intelligences proposes that each person possesses at least seven kinds of intelligence: linguistic, logical-mathematical, musical-rhythmic, visual-spatial, bodily-kinesthetic, interpersonal and intrapersonal. The development of each intelligence within an individual person depends on a great number of factors. A key point is that *each* person possesses a capacity for *all* seven intelligences and the development of each intelligence lies somewhere on the continuum from highly developed to somewhat underdeveloped creating a unique intelligence spectrum for each individual.

For the purpose of identification and discussion, Gardner has separated the intelligences. However, he makes it clear that intelligences are always interacting with each other and that, with the exception of unusual cases, no intelligence exists autonomously in isolation. In an attempt to explain the interdependence of the plurality of intelligences, Gardner (1993) uses the following term:

> ***translation:*** working through a strong intelligence to strengthen a weaker one or expressing a concept from the domain of one intelligence through the medium of another intelligence.

MULTIPLE INTELLIGENCE THEORY AND EDUCATION

Specifically, within the theory of multiple intelligences, it is this concept of *translation* that is most relevant in language learning. In an English language classroom there is much emphasis placed on linguistic intelligence since learning English is, of course, a primarily linguistic endeavor. However, in everyday life, our linguistic intelligence, and thus our language usage, is usually coupled with one or more additional intelligences. For example, a language learner with an underdeveloped linguistic intelligence may have a highly developed mathematical-logical intelligence. In this case, the mathematical-logical intelligence can be a means or medium for communicating the content of the English language lesson. Thus, instead of isolating the intelligences, the student may benefit from an activity that involves the combination of linguistic intelligence and logical-mathematical intelligence providing an opportunity for the intelligences to work together. This interaction more closely resembles how the intelligences naturally function in a more authentic environment

outside the classroom. Therefore, the interaction of the intelligences may serve to heighten both the student's interest in and ease of learning English language subject mater.

Educators who are not familiar with the theory of multiple intelligences might, however, be familiar with the theory of learning styles and might be making a connection as they read this article. While there is, of course, an apparent relationship, the theories remain distinct. Thomas Armstrong (1994) explains the differences:

> Multiple Intelligence theory is a *cognitive* model that seeks to describe how individuals use their intelligences to solve problems and fashion products. Unlike other models that are primarily process oriented, Gardner's approach is particularly geared to how the human mind operated on the *contents* of the world (e.g., objects, persons, certain types of sounds, etc.). A seemingly related theory, the Visual-Auditory-Kinesthetic model, is actually very different from Multiple Intelligence theory in that is a *sensory-channel* model (Multiple Intelligence theory is not specifically tied to the senses; it is possible to be blind and have spatial intelligence or to be deaf and be quite musical). (p. 14)

MULTIPLE INTELLIGENCE THEORY AND ASSESSMENT

In light of the theory of multiple intelligences as well as other current intelligence theories, it may be necessary for educators to reconsider their instructional practices to allow more freedom for the different intelligences. (Sternberg, 1985) In my own personal case, as a language teacher, I have made an effort to encourage multiple ways of conceptualizing and organizing knowledge during classroom activities, prompting students to offer different ways of framing questions and conducting inquiry. This is based on the premise that if students are given the opportunity to frame problems in their own way, they may strengthen their understanding and identification with their own unique means of making sense of the world around them (Hearne & Stone). Allowing students to conceptualize and organize knowledge different ways classroom activities is much less controversial and problematic, however, than allowing students to *demonstrate* their knowledge in different ways by having them take different tests. The problems are objectivity and fairness. But, if instruction reflects the new conceptualizations of intelligence then so too should assessment. Doug Brown (1994) explains:

> These new conceptualizations of intelligence give us both freedom and responsibility in our testing agenda. We are free from *only* using times, discrete-point, analytical tests in measuring language. We are liberated from the tyranny of "objectivity" and its accompanying impersonalness. But we are responsible for *also* tapping into whole language skills, learning processes, and the ability to negotiate meaning. we must therefore test interpersonal, creative, communicative, interactive skills, and in doing so place some trust in our subjectivity, our intuition. (p. 377)

INTERACTIVE VOCABULARY QUIZZES

I decided to experiment by combining Multiple Intelligence theory and vocabulary quizzes. I chose vocabulary quizzes because they are a necessary, yet often a tedious and somewhat repetitious classroom activity focusing solely on linguistic intelligence in a relatively artificial manner (based on the idea that outside of a testing situation, it would be a rare case for a student to be asked to *define* an English word

using merely a linguistic definition). So, I decided to have students take vocabulary quizzes that would allow them to use *translation* by providing choices in how to show mastery/competence (Herman, Aschbacher, and Winters, 1992). Students were provided with the opportunity to "use a stronger intelligence to 'train' or empower a weaker one" (Lazear, 1994, p. 81) when taking vocabulary quizzes.

The students in my 'Intermediate I' English class at Kwansei Gakuin University in Nishinomiya, Japan were the first to take what were to be called *Interactive Vocabulary Quizzes*, interactive meaning between students as well as between intelligences. The twenty women and five men in the class were all second-semester freshmen in the school of Humanities. The class met for an hour and a half three times a week for one semester and would also be together for 'Intermediate II', to occur the following semester.

In the second week of the semester, students studied the Multiple Intelligences theory and were told that we were going to apply it to learning and assessing vocabulary. I tried to emphasize that since this was an English class, we would *always* be nurturing their linguistic intelligence and that we were going to try to combine it with varying other intelligences as occurs naturally outside the classroom. I also explained that we would always be nurturing their interpersonal intelligence in accordance with the notion of language as communication between two or more persons and the consequent importance of *interactive* language learning. Throughout the semester, the students took six quizzes. I designed the first quiz for modeling purposes, and the students designed the remaining four quizzes.

The students self-selected Interactive Vocabulary Teams consisting of four students made up of two pairs, pair A and pair B. Students stayed in the same team throughout the semester. During the 30 minute quiz-taking time, one pair quizzed the other pair, and then vice-versa, a system devised to facilitate interactive language learning. To provide a model, I created the first quiz which consisted of Pair A kinesthetically miming words for the students in Pair B who then had to guess which words were being mimed. For the subsequent quizzes, the pair created quizzes for each other (done in class on the day just prior to the day of the quiz). Having the students create the quizzes was necessary since each pair chose the words on which they would be quizzed (from a regularly maintained vocabulary notebook). Each pair also chose which intelligences would be the means for how they would be quizzed. For example, each individual in Pair A chose 10 words from his/her vocabulary notebook (they are required to enter 20 new words weekly), combined them with his/her partner's 10 words, and together, they wrote the words and their definitions on a teacher-provided handout to be submitted to Pair B. In a small box located on the handout, they also indicated which intelligence would be the main means (to be coupled with linguistic intelligence) for the vocabulary quiz.

Once Pair B received the completed handout, they chose 10 of the 20 words and then began to create a quiz for Pair A. Creating the quiz usually required 20-40 minutes, largely dependent on the structure of the quiz. For instance, in once case, Pair A wrote "visual-spatial" in the Multiple Intelligence box so Pair B designed a test similar to the popular game, "Pictionary" in which the individual students in pair B drew pictorial representations of the words for their partner who then had to guess which word was being drawn. This was not a very time-consuming quiz to create although it *did* take them approximately twenty minutes to conceive of it. On the other hand, an altogether different Pair A in a different vocabulary team wrote "musical-rhythmic" in the Multiple Intelligence box and so Pair B in *their* team wrote an original song incorporating Pair A's 10 vocabulary words into the lyrics. On the day of the quiz, they sang the song omitting the vocabulary words so that Pair A had to fill in the missing vocabulary words on a lyric sheet. In this case, the song writers created the majority of the quiz outside of class.

Before the students began to design quizzes, we generated ideas in class and I provided them with a handout of examples. I strongly encouraged creativity and

originality, however, I spent a significant amount of time monitoring and negotiating the design of the quizzes. Each pair was required to write detailed directions for the quiz on a specific handout so that I could ensure a certain level of congruency.

STUDENT FEEDBACK

At the end of the semester, I asked the students to evaluate the advantages and disadvantages of the Interactive Vocabulary Quizzes. Of the 25 students in the course, 19 claimed that they were "very satisfied" with the experience, four were "satisfied" and two were "unsatisfied".

I asked the students to explain their answers and the following is an excerpt from one student who was "unsatisfied" with the experience:

> *I think we make test by ourselves the level is different from each group.....I think that such a long time for such a test for only 10 words is not worth doing.....I'd like to have a test which is done with whole class with a tense atmosphere or simple paper test.*

The other student who was unsatisfied claimed that he did not make an effort to learn his partner's words, nor vice versa, so they each guessed only their own vocabulary words on the quizzes.

Interestingly, all of the students who claimed to be "satisfied" with the quizzes, nevertheless expressed that the quizzes took a lot of time. One student explained:

> *In the beginning, I sometimes thought it took too much time. I thought it was better that we just memorized the words and our team member told us the meaning of the words and we guessed because it took short time, but we can do such a think in every English classes, so now I think it was good for us to take much time by using various kinds of intelligences.*

This student also mentioned that the individuals in each pair normally answered the words that they chose and not their partner' s words.

Of the 19 students who claimed that they were "very satisfied" with the experience, only one addressed the 'fairness factor'. S/he explained:

> *I think it's very meaningful to have interactive quizzes. Someone may say "it's unfair to have different quizzes in same class", but I don't think so because the points we take (10 or 9 or 8 or 7 or...) is not so important, creating quizzes and thinking is more important to progress our English ability, I think.*

The remaining seventeen students explained that they were "very satisfied" with quizzes because they could "relax and enjoy working together" with their classmates, they could "become more creative" and that they were "better able to memorize words." In fact, one student explained:

> *I felt the quizzes were very interesting because there were many kinds of quizzes such as logical puzzles, art appreciation, mime, and so on. When I was a junior high school and high school student, it was difficult for me to memorize many English words, but now I can memorize a lot of English words with pleasure.*

The final student who was "very satisfied" gave the most comprehensive explanation:

> *The quizzes were very helpful and encouraged me to expand vocabulary in various ways. I tried to expand my vocabulary with only linguistic intelligence. And I*

believed it is difficult to expand vocabulary without (only) linguistic definition. . . . When I was a child, did I look up each word in the dictionary? No, I acquired Japanese in daily life with using multiple intelligences. For example, can we explain a spring (she inserted a drawing of a mechanical spring) or a curve ball in baseball without using gestures or visual aids? I think it's impossible. Multiple Intelligences are essential in expanding vocabulary, I think.

IMPLICATIONS

The feedback from the students convinced me that doing the Interactive Vocabulary Quizzes was a beneficial form of assessment so I have decided to continue the quizzes for this same class next semester. However, I will be more attentive to the amount of class time used for preparing the exams and I will re-structure the design so that both individuals in each pair are held more accountable for learning each other's vocabulary words. I am pleased with the outcome of the quizzes and would encourage other instructors to try implementing alternative methods of assessment which take currently developing theories of intelligence into account.

REFERENCES

Armstrong, T. (1994). *Multiple Intelligences in the Classroom.* Alexandria, VA: The Association for Supervision and Curriculum Development.

Brown, D. (1994). *Teaching by Principles: Interactive Language Teaching Methodology.* New York: Prentice Hall Regents.

Campbell, B., Campbell, L., & Dickinson, D. (1992). *Teaching and Learning Multiple Intelligences.* Seattle: New Horizons for Learning.

Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice.* New York: Basic Books.

Hearne, & Stone. (1995). Multiple Intelligences and Underachievement: Lessons from Individuals with Learning Disabilities. *Journal of Learning Disabilities, 28,* 339-448.

Herman, J., Aschbacher, P., and Winters, L. (1992). *A Practical Guide to Alternative Assessment.* Alexandria, VA. Association for Supervision and Curriculum Development.

Lazear, D. (1994). *Multiple Intelligence Approaches to Assessment: Solving the Assessment Conundrum.* Tuscon, Ariz.: Zephyr Press.

Sternberg, R.J. (1985). *Beyond IQ.* Cambridge, England: Cambridge University Press.

Part 3

Current Assessment Practices at KGU

Chapter 7

What Do We Do Now?
A Survey of Current Assessment Practices at KGU

PRISCILLA BUTLER

Though multiple examples of survey research on the attitudes and perspectives of Japanese college students studying English exist (Hadley & Hadley, 1996 and Shimizu, 1995, for example), a survey of research listed on English language databases turned up no such examples of surveys oriented toward faculty teaching English at Japanese colleges or universities. This paucity of information about teaching and assessment practices, as well as university specific and wider domain challenges English teaching faculty at these institutions face, limits the degree to which foreign faculty can know, much less understand, the teaching context in which they are employed.

As we[1] began our own research project on alternative assessment, a central question members had focused on current practice among other faculty members. As Wordell (1993) notes, many foreign instructors have somewhat of a periphery relationship to Japanese and other faculty with attendant limited opportunities of talking with them about their classes or getting to know the deeper structure--educational, social, and political--of Japanese university life. Because one of our central purposes in working on this project was not only to describe some of our own emerging alternative assessment practices but also to understand better the context in which we are using these assessments, we felt it incumbent on us to make an effort to find out more about the teaching situations of other faculty, especially Japanese faculty members. The resulting survey research project was our attempt to best meet this need.

METHOD

Before developing the research instrument, seven exploratory interviews were held with faculty members who agreed to talk with us about their teaching and assessment practices. Open-ended research questions were developed as a rough guide for the interview, but, in keeping with Seidman (1991), a primary purpose was to allow interviewees to control the structure of the interviews in order to elicit a more emic (speaker-oriented) perspective.

The seven faculty members included four full-time and three part-time instructors from four different schools/departments. Though convenience sampling was used in this initial stage of the research, an attempt was made to speak with a cross-section of faculty who might represent different views. Interviews were recorded, but interviewees were guaranteed confidentiality and anonymity in order to encourage openness. Complete transcripts were made of all seven interviews. Those interviews held in Japanese were translated into English, and survey group members read these transcripts, analyzing themes and issues from which they derived an initial set of survey questions.

[1] Generally, the Kwansei Gakuin University Alternative Assessment Research Group; More specifically, those group members who developed the survey instrument: Priscilla Butler, Shalle Leeming, Maricel Santos, and John Chapman-Rienstra.

The first version of the survey was pretested by selected faculty members taking part in the research committee and translated into Japanese and backtranslated into English to ensure an exact match of both wording and meaning. The Japanese version was further checked by two additional native Japanese speakers who confirmed its correctness. After additional feedback from research group members was solicited, a revised version was produced and used in a pilot study. Methodological procedures for initiating the pilot study were in keeping with Babbie (1990) and Goldenberg (1992).

Based on a sampling frame of a list of all faculty members obtained directly from each school within the university--excluding committee members as well as instructors on sabbatical or otherwise not teaching that semester--a systematic sampling of the population was completed. The original sample consisted of 20% of the entire population of 197 teachers. Surveys were sent with a personally addressed explanatory letter and a fully addressed campus mail envelope to ensure a higher rate of return. Forty-four percent of faculty members returned the pilot survey.

The Final Survey

Based on this return rate, the committee decided to send the final version of the survey to all remaining faculty members, rather than a smaller sample. Assuming a similar rate of return, it was predicted that approximately 70 of the final surveys would be completed. To better ensure this rate of return, in addition to supplying a personal letter of explanation and a pre-addressed envelope, a follow-up reminder letter was also sent two weeks prior to the deadline. Before the final surveys were sent out, minor modifications were made to the survey, based on the responses received. Frequently omitted questions were framed in boxes and put in bold. Questions which had ambiguous or qualified responses were reworded to increase clarity.[2] Both closed and open-ended questions were included to maximize the variety of ways in which respondents could describe their teaching and assessment practices. Teachers were not asked to indicate school or departmental affiliation because it was felt this might cause undue concern about any attempts to note differences among departmental practices.

Surveys were sent via campus mail to a total of 158 English teachers at Kwanset Gakuin University, the entire remaining population of English teachers--excluding committee members and those who had previously been sent the pilot survey. Seventy-one (43%) returned the survey. As Babbie (1990) points out, a return rate of 60 percent is considered good; however, typical mail delivered surveys usually have substantially smaller return rates (p. 182). Of course, survey research often relies on relatively small sample sizes, assuming statistical probability will limit the degree of sampling bias that occurs. The final version of this survey; however, was mailed to 80% of the population being surveyed, a very large sample size, thus ensuring a reduced chance of sampling bias.

The survey results were analyzed qualitatively for recurring patterns, as outlined in Goldenberg (1992: 299). Though descriptive statistics were computed for selected numerical patterns that emerged, results are recorded in narrative summary as well.

RESULTS

Initial analyses were conducted with two unifying frames of reference: 1) class type (reading, writing, speaking, listening) and 2) status of employment (full-time or part-time). It was hypothesized that these two groupings might elicit distinct patterns of response.

In setting off the domain of class type, we instructed the participants to choose one and only one class that they most often teach and apply all survey answers to that

[2] See Appendix A for the final version of the survey.

particular class type. In doing so, we sought to eliminate the burden of time each participant[3] would need to complete the survey and to focus the participants' attention on one unified class rather than taking examples from multiple classes perhaps very different from each other. We also realized that, given a choice, teachers might choose to describe a class that would reflect most flatteringly on themselves; we hoped to counteract this by explicitly directing teachers to write about the class most often taught. Nonetheless, as in all survey research, we recognize the possibility that teachers themselves may have self-selected particular kinds of classes.

We further asked teachers, in addition to marking a specific class title, to check off which of the four skills they felt they focus on in that class. If they focus on multiple skills but feel one is *primarily* emphasized, they were instructed to underline that main skill area.

Table 1: Percentage of English Classes Oriented Around Specific Skills

Skill Area(s)		Percentage of Classes (# of surveys)
Reading	(Main or Only Skill)	**44%** (31)
Reading	(An aspect of the class)	66% (46)
Writing	(Main or Only Skill)	**17%** (12)
Writing	(An aspect of the class)	36% (25)
Listening	(Main or Only Skill)	**11%** (8)
Listening	(An aspect of the class)	47% (33)
Speaking	(Main or Only Skill)	**7%*** (5)
Speaking	(An aspect of the class)	29% (20)

*Totals do not equal 100% because participants could select as many or as few skill areas as desired. Some participants did not select a main skill.

Fully 44% of participants reported "reading" as being the main or only focus of the class they had selected. When "reading" as an additional or equally important skill focus of a class was factored in, another 17% of respondents focused on this skill for a total of 66% focusing on reading in the classes they described. Moreover, the distribution of classes focusing on other skills was too small to demonstrate a consistent pattern. The one skill area which a number of respondents selected, "writing," only amounted to 12 of the 71 surveys, when counted as the main or only skill. Furthermore, among these twelve focusing on writing, eight (or 66%) reported that they used translation as a means of assessment.[4] This definition of writing does not fit the Western conception of it but seemed to more closely resemble the way reading classes are defined in Japan, as translation is the most popular assessment technique reported among reading teachers. The results of this analysis of the surveys as divided by skills reveals that those who took part in our survey focused on classes which are primarily

[3] In keeping with the a qualitative epistemological perspective, the term "participants" is used here deliberatively instead of the more often used "subjects," which is a term associated with the quantitative research mode. The words reflect something of the researcher's stance toward what is being studied. "Subjects" implies that people are the object of study, but the primary "do-er" is the researcher. "Participants" implies that the researcher recognizes the collaborative role of those who choose to take part in a study. For further discussion of this issue, see Ely (1991).

[4] This pattern will be examined more explicitly later in the paper. For a more complete analysis of this issue as it relates to the teaching of writing; however, please refer to chapter 4.

reading oriented. The uniformity of response prevented further generative analysis by this initial grouping.

UNDERSTANDING OF CLASS GOALS

One interesting pattern that emerged from a coding of class type as it related to the skill the teacher emphasized in class was that, though there was some agreement across classes, there was quite a bit of variation even with a group of classes which share the same class title. Though this may partly be due to different departmental interpretations of whatever rough guidelines exist for courses, the diversity of interpretations of course focus points to a lack of shared purpose in certain courses. The course called "I-otsu" is the best example of this.

Table 2: Skills Emphasized in Particular Classes

Course Title	# of sections	Primary Skill(s) Taught
I-ko	11	R, R, R, R, Rw*, RL, RLS, RLS, RL, RL, LS
II-ko	13	R, R, R, R, R, R, R, R, RW, RLW, RL, RL, L
I-otsu	10	W, W, W, W, WL, LRW, LRW, LS, SL, L
II-otsu	5	R, R, R, RL, WR
Eikaiwa (I, II)	5	S, LS, LSw, SL, SLRW
English A	5	R, R, R, R, W
English B	2	LRS, WS
English C	2	R, WR
Writing I, II; Comp. II	3	W, W, W
English Expression	3	LWS, LW, LSRW
English I	2	LS, LSRW
English II	2	R, RW
Bridge to Human Ecology	2	LR, LR
Ad. English A	1	LSR
Eigo Hyogen Enshu	1	SL
IM English	1	SLRW
Speaking I	1	SL
other	1	R

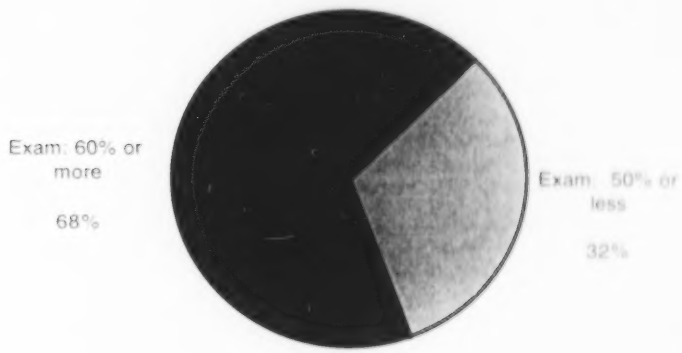
(N=70; This section was left blank on the 71st survey.)
*Skills indicated in large type are the primary focus of class; those indicated in smaller type are still taught but to a lesser degree.

Though five of the classes are writing focused, the remaining five fall into a number of different categories including: a purely listening focused class; a speaking focused class with some listening; a class focusing on both speaking and listening; one focusing on listening with a reading and writing component; and a fifth focusing on shared skills of listening, reading, and writing. These results show that there is little agreement among faculty members about the purpose of individual courses which share the same title.

THREE GROUPS OF TEACHERS AND EXAM ORIENTATION

A second analytical frame, employment status, also did not elicit differentiated patterns of response between the two divisions: part-time teachers and full-time teachers.[5] The full range of responses were exhibited within both groups. A third category of division was then devised distinguishing Japanese and foreign teachers. This division proved fruitful as the two groups exhibited markedly different types of responses; however the categories seemed again too broad. More extensive analysis of the data revealed an additional division: teachers who focus primarily on the final exam (defined as weighting the final at 60% of the grade or higher) and those who do not (weighting the final exam at 50% or less).

Table 3
Final Exam as a Percentage of the Final Grade: Japanese Teachers



This further categorization was arrived at after looking at each potential grouping and cut-off point, including having the final exam as 70% or 80%, among other divisions. Careful analysis of the qualitative data, though, revealed that some teachers reported being part of a group, department, or program which required an examination worth 50% of the course grade. Although not all teachers reporting a final exam worth 50% of the grade also volunteered information about whether this was an outside expectation or not, that at least some teachers explained that this was an agreed upon criterion suggested that teachers who had final exams worth up to 50% of the grade might not be doing so voluntarily, in the strictest sense of the word. Thus, it was decided that 50% would be the cut-off point for separating the two groups.

[5] Though three divisions were made on the survey (full-time, part-time, and full-time: contractural basis), the two full time distinctions were collapsed in the analysis itself.

Interestingly, no one reported having any final exams worth something between 51 and 59% (55%, for example).

Table 4
Final Exam as a Percentage of the Final Grade

(Foreign Teachers)

Final Exam %	# of Teachers	% of F. Ts
0% (No final)	11	70%*
15%	1	6%
30%	1	6%
40%	2	12%
50%	1	6%

(N=16)

* Percentages have been rounded.)

Table 5
Final Exam as a Percentage of the Final Grade

(Japanese Teachers)

Final Exam %	# of Teachers	% of J. Ts
0% (No final)	3	5%*
20%	1	2%
25%	1	2
30%	1	2
40%	2	4
45%	1	2
50%	8	15
60%	8	15
70%	12	22
80%	11	20
90%	2	4
100%	4	8

(N=54; The 55th survey was left blank.)

* Percentages do not equal 100 due to rounding.)

The percentage of weight assigned to the final exam was considered important because the final exam tends to indicate something of the way a teacher may be thinking about how knowledge is transmitted or acquired. Teachers who place a strong emphasis on the final exam may view assessment's primary role as being a means of evaluating students in order to assign grades or let a student know how his performance on a test compares to others. In this exam focused system, the primary emphasis would be on maximizing the amount of information or knowledge a student can demonstrate

remembering. Teachers may believe that their job is to transmit as much knowledge to students as possible. This has been called the common sense view of education (Mayher, 1990).

Teachers who choose to assess students in multiple instances in multiple ways may tend to adhere to a different concept of education: one that assumes that learning may not necessarily be an accumulation of information rather than an engagement with a variety of concepts. This view would emphasize more student involvement. Learning would not necessarily be viewed as something that only takes place when the teacher passes on knowledge/information to students but also when students interact with each other, the ideas themselves, or even people outside of the classroom. The first view of education would tend to present knowledge as an objective set of facts to be adopted, while the second would tend to view knowledge as a constructed artifact which relies on the active participation of learners in its construction. In the first system, lecture mode and end-of-term assessment would be emphasized. In the second system, participatory interaction and ongoing assessment would be emphasized.[6]

EXAM ORIENTATION AND DISCLOSURE OF GRADING CRITERIA

Of course, a single teacher may evidence examples of both modes in her or his belief system; still, the divisions are useful for looking at different stances toward the educational process. In order to check my hypothesis that these two groups of teachers might look at the learning/assessment process differently, I compared the answers participants gave about how much they inform students of their assessment practice. The teachers had been asked to select from four options: 1) that they explain both the factors counted toward final grades and the percentage of the grade assigned to each factor, 2) that they explain only the factors to be graded, 3) that they do not explain how they grade, or 4) an option to fill in any other answer not presented. I hypothesized that the group of teachers who weighted final exams very heavily would also be more likely not to inform their students of all of the grading details: in other words, to choose the second, third or fourth selection over the first.

Table 6: Percentage of Teachers who Explain Grading Criteria, by Group

3 Groups	explains all	explains only factors	does not explain
Foreign Ts (N=16)	80%	15%	5%
Multi-Assessment oriented Japanese Ts (N=18)	80%	13%	7%
Exam-oriented Japanese Ts (N=36)	50%	42%	8%

(N=70; On one survey, this section was left blank.)

Table 6 shows that the group of foreign English teachers and the group of Japanese teachers who weight the final exam at 50% or less are almost identical in their patterns of response to this question. Eighty percent of both groups fully disclose their grading criteria, at least for the major divisions of grades asked about in this survey. In

[6] For a more extended discussion of these two views of knowledge, the roles of teachers and learners, and the role of education, please see chapters 1 and 2.

comparison, of the group of teachers weighting the final exam 60% or higher, only 50%--or 30% fewer--teachers explained grading criteria completely to their students.

EXAM ORIENTATION AND TYPES OF ASSESSMENTS

In order to check whether these analyses were consistent with the rest of the items in the survey, a second stage of analysis focused on the content of the open ended assessment descriptors. Participants had been asked to describe the types of assessments they do in their class and to explain how they evaluate each type. All of the open-ended written responses were coded, and frequency counts were made of each type of assessment given as a part of the final examination. Fully 65% of the exam oriented teachers used translation as a part of their final exam, compared with only 33% of the group of Japanese teachers which emphasized multiple assessments. None (0%) of the foreign teachers used translation as an assessment tool.

Table 7: Five Most Common Types of Assessments on Final Exam, by group

Foreign Teachers	Multi-Assessment J. Ts	Exam oriented J. Ts
No Final (69%)	Translation* (33%)	Translation (65%)
Group Test (12%)	No Final (22%)	Summary/Comprehension (20%)
Oral Debate or Presentation (12%)	Summary/Comprehen-sion (22%)	Composition (20%)
Oral Interview (8%)	Composition (5%)	Cloze Passage (17%)
---	Multiple Choice (6%)	Fill in the blank (11%)

*Includes translation of words only as well as phrases and passages.

Translation has been considered a well know, time honored practice in Japan (Hansen, 1985) and, as such, tends to typify a more traditional view of education. Though the group of Japanese teachers who favor multiple assessments still used many other traditional assessment approaches, they were less likely to use the most traditional or conservative of these practices, such as transposing word order (taking a stream of words and putting them in the correct English word order) or reproducing a composition from memory as it appears in a textbook: both practices cited by a few of the exam oriented teachers. Further, they were less likely to give an exam (22% did not compared with 0% in the exam oriented group), and they were more likely to have more than three different forms of assessment in each class.

These results support the hypothesis that the percentage a teacher assigns to the final exam as well as how much she or he discloses about grading criteria are linked with the types of exam questions she or he asks. In other words, these two rough divisions seem to represent groups of teachers who have certain different educational beliefs. It is important to note, however, that there are twice as many teachers in this study who fall into the exam oriented category. Those who try even slightly less traditional assessment practices are distinctly in the minority, making up only one third of the Japanese teacher sample.

Both groups, however, contrast sharply with the practices of foreign teachers, primarily because 11 of the 16 teachers reported not giving a final exam in their classes. In this case, the choice may be partially linked with the type of class being taught, six of which were explicitly speech-oriented classes, including various sections of *Eikaiwa* (English conversation) and English I or Speaking I, oral communications oriented courses. Of those who gave final exams, though, two gave tests that had been developed by a group of teachers and were shared by all teachers responsible for that type of class; two asked students to do a group oral debate or presentation for their final exam, and one gave a one-on-one interview test as the final examination. It should be noted that no (0%) of foreign teachers required a final examination worth more than 50% of the grade, and the two who did have a 50% weighted final exam indicated that this was a shared exam agreed upon within a group of teachers, not something they individually controlled.

CHALLENGES AND PROPOSED CHANGES

No such patterns of difference were found among the results of the last two open-ended questions on the survey: 1) What do you think is the biggest problem or challenge that university English teachers in Japan face in terms of assessing student progress? and 2) If you could change anything about your current teaching and/or assessment situation, what would you like to change? Similar responses were recorded across all three groups.

Participants viewed the biggest challenges faced in Japan as being: large class sizes, lack of student motivation, teachers' lack of enthusiasm, the infrequency/lack of time the class meets, and the discrepancy of level among students. The changes at Kwansei Gakuin University that were most desired include: reducing the size of classes, increasing the frequency of meeting/the amount of time in class, and dividing the students by level. Though both questions elicited a number of other responses, no single response produced reached a frequency as high as five. The number and diversity of comments make a complete listing here unwieldy. In addition, less frequent responses were thought to represent individual rather than shared views.

SUMMARY OF RESULTS

Survey results suggest that there are at least three major divisions of teachers at this university, and these divisions do not necessarily fall along lines one would expect such as by the type of classes for which teachers are responsible or their employment status as either part-time or full-time employees. Rather, data analyses indicate that they seem to be divided into roughly three groups according to their assessment practices, as indicated by the percentage of weight they assign to final exams, their willingness to disclose grading criteria to students, and the types of questions they include on their final examinations. The three groups are: 1) exam oriented Japanese teachers, 2) multiple assessment oriented Japanese teachers, and 3) foreign teachers who are less inclined to give any final examination at all. By far, the largest group is the exam oriented Japanese teachers who number twice as many as the multiple assessment oriented Japanese teachers.

Other findings include that there seems not to be much agreement about the common aims of certain courses, even those that share the same class title; that regardless of class title, Japanese teachers are extremely likely to use translation as a method of assessing the class; that reading is reported as being the most focused on skill by Japanese teachers; and that teachers most hope for reduced class sizes, more contact hours with each class.

DISCUSSION

Though this is a survey about assessment, it is also intrinsically a survey about curriculum as well. Assessment is not divorced from curriculum; rather it is a central component of what happens in the classroom as assessments very often reflect or even determine the kinds of activities that occur and the ways knowledge is discussed. Before turning to specifics of assessment practice, then, it makes sense to first look at issues of curriculum along with assessment within the context of the social, cultural, and political constraints that influence both.

Status of Different Skills: Issues of Curriculum and Purpose

In the current university English curriculum, it may not come as any surprise that reading is the skill that most Japanese teachers report focusing on in the classes they teach most often. Japanese English teachers at the tertiary level are generally experts in some aspect of literature or linguistics rather than language or pedagogy. As Carrell (1988) notes, for many EFL students the desire to read is often a central reason for studying the language because most students are more likely to encounter written texts in their future personal and professional lives than to have multiple contacts with English speakers.

Survey research on Japanese learners' actual interests, however, suggests that this group of students actually is more interested in learning how to speak and listen to English than to read or write it (Kobayashi, Redekop, & Porter, 1992; Widdows & Voller, 1991). This apparent division between what many teachers currently offer and what many students profess to want to learn may also have more to do with how the skill of reading is defined.

Definitions of Reading

Educators hold different views about the origin of current reading pedagogy in Japan. Hansen (1985), for example, asserts that the current translation orientation of reading classes derives from the Meiji period when the role of universities was primarily to "introduce Western technology through translations" (p. 151). Bamford (1993) describes current reading pedagogy as fitting in with a traditional grammar translation method of language teaching:

> In a university grammar-translation class, the words of an English sentence are each given a Japanese equivalent with the aid of a bilingual dictionary. The Japanese words are then reordered to provide the closest approximation of the English original. Students are called on in class to give their rendition in Japanese after which the teacher gives a model translation, and, in this way, two or three pages of text are covered each class. (p. 64)

Law proposes a different view of reading pedagogy and its focus on translation. Instead of describing it as grammar translation with its roots in 19th century European models of language instruction, he traces it back to a method called *yakudoku*, or translation reading, which he asserts developed in Japan several centuries ago for "decoding ancient Chinese texts" (p. 215). He explains one difference between traditional grammar translation and yakudoku as being one of purpose, in that yakudoku "reflects the classical assumptions in that it focuses more on understanding the valued contents of the translated text than on mastering the codes of the language itself, and in that it is concerned predominantly with the one-way transmission of ideas from the foreign language" (p. 215).

What each of these treatments of reading pedagogy implicitly confirms, however, is that such pedagogical systems do not fit current Western notions of what it means to read. As Bamford asserts, "If you are going to teach a reading course at a university, you should know that most of your students probably have not read more than a few words of English in their lives" (p. 63). The key underlying assumption here is that reading has more to do with "rapid, automatic" processing of texts--including interacting with texts as readers bring their own background knowledge and experiences to bear on what is being read--than with being able to make exact translations of difficult English texts. In fact, many Western-trained reading instructors would not necessarily associate reading with translation at all, and certainly not on a sentence by sentence level.

Diverging Views of English Education

This apparent fundamental difference in defining what it means to read also suggests a difference in the purposes of English education. Law (1995) proposes that English education has traditionally served one or more of three major functions: 1) mental training, much as Latin has traditionally been taught in some Western schools as a kind of mental calisthenics; 2) reflection of Japanese itself--the otherness of English helps clarify what Japanese language is, a kind of reverse first language training; and 3) a means of social control--if learners are asked to remember a set of arbitrary rules without any apparent reason for doing so or principle behind such learning, and they perform this memorization well, in the process they learn unquestioning obedience and the value of hard work or effort.

Though Law takes a decidedly political view of the function of translation orientation in English reading classes, his main point is still something which should be looked at more closely: that is, if the reading that students do does not actually serve the purpose of helping students more easily and willingly read English materials, than perhaps the purpose of these classes is an entirely different one. Perhaps it has less to do with the use of language and more to do with the subjectification of language--that is, seeing language not as a tool for actual communication of ideas, whether in the written or spoken form, but as a complex system of rules waiting to be mastered. If English classes are viewed in this way, a status split between the oral/aural skills and the text-oriented ones makes sense.

In nearly inverse proportions to the degree that reading classes are taught by Japanese faculty members, speaking and listening oriented classes tend to be the domain of foreign teachers. After all, the particular skills and abilities that these two groups have might indicate a similar division. Many universities do hire foreign instructors primarily to teach oral communications courses. Nonetheless, as Law (1995) points out, such divisions may help to further the ideological split between the way English is viewed in Japan. He suggests that the separating of "eigo (English language) and eikaiwa (English conversation)" further distinguishes the foreign from the familiar in that English language studies are viewed as another content subject, much like history, while English communications courses tend to be viewed as somewhat frivolous, sometimes fun, but definitely not serious objects of scholarship or education. He asserts that, "Unless and until we see greater numbers of native-speaker teachers involved in teaching reading skills, for example, and of non-native teachers seeing it as a primary duty to teach oral skills, it will be difficult to convince students that all are engaged in the same enterprise, and that communication skills are not marginal aspects of language learning" (p. 222).

Results from a survey of 1,088 students at eight Japanese women's colleges suggest that students already view their foreign and Japanese English teachers markedly differently, with 28% reporting that intelligence is an important quality in Japanese English teachers compared with only 4% stating it is an important attribute for foreign teachers. Similarly, 34% of these students rated knowledge of the subject area as

important for Japanese teachers, whereas, only 7% thought this was important for foreign teachers. Conversely, students' top two characterizations of their Japanese English teachers' classes are as being "very formal" (80%) and "Make me sleepy" (76%), though the top two characterizations for classes with foreign teachers are "cheerful" (75%) and "fun" (71%) (Shimizu, 1995).

I have previously studied this tendency of EFL/ESL learners to dichotomize fun and serious study as well as to label one or the other more or less important (Butler, 1995). Language learners tend to view language classes not taught in the modes they have become accustomed as being less serious than traditionally taught classes, at least initially, because the idea that learning can be an engaging process is so alien to many of their school experiences. This dichotomization is furthered when the types of classes which might warrant the most interaction are limited to foreign teachers. That these teachers also tend to be part-time employees and in lower status positions in the university further decreases the external cues that such classes might be real learning environments.

I do not point out this discrepancy to in any way negate the fact that there are less than responsible foreign English teachers whose classes do not appear to have a well-organized focus for learning. This is true of at least some teachers from whatever nationality. However, I would agree with Law's suggestion that in order for the variety of English classes offered to be viewed as similarly important, useful, or worthwhile, a reconceptualization of what it means to learn English is in order. Issues of curriculum, and, subsequently, of assessment, however, can not be divorced from larger questions, such as what it means to be a university teacher and what the enterprise of university education is all about.

Views of University English Teachers in Japan

Christopher (1983), Kelly and Adachi (1993), and Hansen (1985) all assert that the primary role of a university teacher in Japan has historically been, and in many cases, even now is, that of "scholar" rather than "educator." As Kelly and Adachi note, "This 'teacher as scholar' perspective leads many Japanese professors to consider teaching the least important of their activities" (162). They see a division, however, even among faculty in those who support a purely traditional concept of English education and those who support more progressive educational modes. They suggest that the traditional prioritizing of scholarship and research over pedagogical practice, nonetheless, is in direct conflict with the values and expectations of some less traditional faculty members who value both and may even hold less traditional views of what a good teacher does.

Such divergent views, they propose, underlie many current university conflicts, such as "use of funds for research versus teaching aids; hiring instructors on the basis of scholarship versus experience; [and] centering a writing curriculum on translation versus composition" (p. 163). This divergence in viewpoints even among Japanese university faculty is consistent with the tentative findings of this survey that Japanese faculty members, at least in terms of what their assessments seem to say about them, do not form a unified group which shares entirely similar thoughts on what it means to assess, much less teach, English.

Of course, it would be foolhardy to suggest that any group as diverse as an entire faculty of English language educators would share similar views of teaching and assessment. Still, it does seem worthwhile to point out that the divergence in assessment, and therefore, of curriculum, is not just an issue of how foreign instructors differ from their Japanese counterparts but about diverging views among Japanese faculty members themselves.

ARE PART-TIME TEACHERS MORE INTERESTED IN EDUCATION?

I noted near the beginning of this report that the initial analyses of data under the categories of full-time and part-time employment did not prove to be fruitful in charting patterns of response. One interesting point that did derive from this initial categorization scheme, however, is that part-time teachers had a substantially higher return rate of surveys than full-time teachers. When looked at as a group of those sampled, 46% of part-time teachers returned the surveys, compared with 38% of full-time teachers, an 8% percent difference. This would seem counter-intuitive in that part-time teachers usually come to campus one or two days a week and might be more likely to miss the survey, forget about it, or mislay it. Furthermore, part-time teachers might be expected to have less of an investment in giving such information about their teaching because they may not have a protracted or extended relationship with the university.

Full-time teachers, on the other hand, are usually on campus at least three to four days a week, have an office where they can keep school-related documents centrally located on campus, and, due to their continuing employment at the university, might be expected to have a greater investment in sharing and finding out about their own and colleagues' teaching practices. On the other hand, of course, full-time teachers might also have a greater range of demands on their time, such as committee meetings and research obligations; nonetheless, it seems plausible that they might be more rather than less willing to give colleagues information about how they assess their classes.

That the survey results proved otherwise led me to re-examine the open-ended answers part-time versus full-time teachers gave to the final questions about challenges at Japanese universities and desired changes at Kwansei Gakuin University. A minor pattern among part-time teachers responses emerged in which these teachers proposed a broader base of communication among faculty members about their teaching and assessment practice. Proposals included the following:

- *Teachers should make their way of teaching and evaluation public and mutually make productive comments. It should also be a rule for all teachers to submit their assessment criteria to the school and make this information available to the students.*

- *A horizontal network between teachers is necessary for exchanging opinions, talking about how to choose textbooks, and forming a reading, listening, speaking, and composition network.*

- *It would be good if Japanese instructors increased their level of oral communication and connect their class with that of a native speaker.*

- *Teachers should exchange opinions and mutually observe each others' classes more. Some teachers are trying to keep their own area and their situation is, "The frog in the well knows nothing of the great ocean."*

Though only four teachers wrote in such comments, none of the full-time teachers suggested similar changes toward more collegiality and pedagogical focus. Perhaps this is reflective of the scholarship versus pedagogy orientation seen in some universities.

FOCUSING ON PEDAGOGY: WHEN AND HOW

This is not to suggest that either full-time teachers or part-time teachers as a group care more about their teaching or assessment practice than any other group; rather it is

to point out that some faculty members perceive there to be limited opportunities to talk about pedagogy within the academic community. The results showing the diverse ways in which teachers teach the same class with the same title that was discussed earlier seem to support this idea that there may not be adequate opportunities or perhaps even interest in discussing teaching at the university level.

As I point this out, however, I would also like to recognize that the university is a community where teachers have a substantial amount of academic freedom (Terasaki, 1995). This should be respected. In my experience, teachers in general and university teachers more specifically dislike being told what to do by others. Teaching is one of the few places where employees get to be their own bosses, so to speak. That teachers are allowed this freedom also means that they have more opportunities to use their creativity, their originality, their distinct skills and areas of knowledge. This can only be a good thing.

In discussing alternative assessment, or any other educational innovation, for that matter, then, there may be a tendency to push a little too hard in the "new" direction. The results of this survey and the chapters in this volume are not intended as a prodding stick; they are rather envisioned as an invitation for teachers to look more closely at their classrooms, to think about how they view English education within the university context in Japan, and to think about how their vision of education meshes with the assessments they use. It would be a mistake for any one person to try to impose a vision of education on another person, and it would be equally mistaken to demand that a colleague begin to disclose his or her practice. If discussions about education are to be fruitful, it must be recognized that colleagues have the right not to join in, or to decide the contexts when they might feel like discussing such issues. Experience with the interviews that started this survey project suggests that faculty members are rarely asked what they do in a sincere way, that is, without an ulterior motive of trying to convert the other colleague to believe the same thing the asker believes. Proselytizing masked as conversations about curriculum and assessment advantage no one.

For discussions to be productive, too, I believe, it is necessary to recognize the institutional constraints that each teacher faces. That English teachers throughout the university, regardless of nationality or employment status, recognized large class size, the lack of time meeting with each class, and the diversity of levels among students in a single class as being sources of frustration points to a need for institutional change. Class size averaged 40 students or more. It would indeed be difficult to use alternative assessments in these classes. As several teachers observed, multiple choice, easily gradable exams are about the only assessment possibility when overwhelmed with such large groups of students. Clearly, institutional recognition of this recurring problem is necessary. Still, even with large class sizes, many teachers did report doing multiple assessments in a variety of ways.

CONCLUSION: RE-THINKING ASSESSMENT, RE-THINKING WHAT IT MEANS TO "KNOW"

I began the introduction to this volume with the recognition of the limitations we, as foreign instructors, have in terms of being able or even worthy of commenting on the systems which we have been invited to join for a period of time. I am also extremely aware of the dangers of trying to view the inside from the outside and thinking one sees all. Our experiences shape what we see. Mine have surely done so here as well. In reporting the units of analysis I used to look at the survey results and discussing the issues evoked by these results, I am also necessarily reflecting my own vision of reality. The tradition of qualitative research recognizes such interpretive lenses as inevitable.

In the end, it is the teachers themselves who must judge what this information means for the way they think about their individual classrooms and methods of assessment as well as the larger social, cultural, and institutional contexts in which they work. The survey itself and the discussion of it has been intended as a tool for reflection. It seems fitting then, to end with one of the more evocative reflective devices the English language has to offer: a poem. In thinking about the assessments we use, it would serve us well to remember how we define what it means to "know," "understand," or "learn."

FROM MORNING-GLORY TO PETERSBURG
(The World Book, 1928)

"Organized knowledge in story and picture"
confronts through dusty glass
an eye grown dubious.
I can recall when knowledge still was pure,
not contradictory, pleasurable
as cutting out a paper doll.
You opened up a book and there it was:
everything just as promised, from
Kurdistan to Mormons, Gum
Arabic to Kumquat, neither more nor less.
Facts could be kept separate
by a convention; that was what
made childhood possible. Now knowledge finds me out;
in all its risible untidiness
it traces me to each address,
dragging in things I never thought about.
I don't invite what facts can be held at arm's length; a family
of jeering irresponsibilities always
comes along gypsy-style
and there you have them all
forever on your hands. It never pays.
If I could still extrapolate
the morning-glory on the gate
from Petersburg in history--but it's too late.

Adrienne Rich (1984, p. 30)

ACKNOWLEDGMENTS

I wish to thank my colleagues for their help and support throughout this research project: especially Shalle Leeming, Maricel Santos, and John Chapman-Rienstra, all of whom worked collaboratively with me to develop the survey instrument and who also provided important feedback throughout the research process. I also would like to thank Assistant Professor Shinji Kimura who provided much valuable advice on the construction of the survey as well as feedback on its multiple revisions.

The insights of Professor Toshio Koyama, Project Director, also proved to be extremely valuable resources during the length of the project. In this capacity as well, all of the members of the Alternative Assessment Research Project Committee proved well-informed, interested, and helpful. As well, I would like to express my sincere thanks to the many faculty members who completed the survey or otherwise

volunteered to take part in the study. They gave of their time at a very busy period. Finally, I am very grateful for the generous funding provided by Kwansei Gakuin University, without which this research project would not have been possible.

REFERENCES

Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth Publishing Company.

Bamford, J. (1993). Beyond grammar translation: Teaching students to really read. In P. Wadden (Ed.), *A Handbook for Teaching English at Japanese Colleges and Universities* (pp. 63-72). Oxford: Oxford University Press.

Butler, P. (1995). Looking in and seeing out: A qualitative study of an adult ESL classroom. Unpublished doctoral dissertation, New York University, New York.

Carrell, P. L. (1988). Introduction: Interactive appraoches to second language reading. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 1-7). Cambridge: Cambridge University Press.

Christopher, R. C. (1983). *The Japanese Mind.* New York: Fawcett Columbine.

Ely, M., Anzul, M, Friedman, T, Garner, D., & Steinmetz, A. M. (1991). *Doing qualitative research: Circles within circles.* London: The Falmer Press.

Goldenberg, S. (1992). *Thinking Methodologically.* New York: Harper Collins.

Hadley, G., & Hadley, H. Y. (1996). The cutlute of learning and the good teacher in Japan: An analysis of student views. *The Language Teacher, 20* (9), 53-55.

Hansen, H. E. (1985). English education in Japanese universities and its social context. In C. B. Wordell (Ed.), *A guide to teaching English in Japan* (pp. 145-170). Tokyo: The Japan Times.

Kelly, C., & Adachi, N. (1993). The chrysanthemum maze: Your Japanese colleagues. In P. Wadden (Ed.), *A Handbook for Teaching English at Japanese Colleges and Universities* (pp. 156-171). Oxford: Oxford University Press.

Kobayashi, S., Redekop, B., & Porter, R. (1992). Motivation of college English students. *The Language Teacher, 16* (1), 7-9, 15.

Law, G. (1995). Ideologies of English language education in Japan. *JALT Journal, 17* (2), 213-224.

Mayher, J. S. (1990). *Uncommon sense: Theoretical practice in language education.* Portsmouth, NH: Boynton Cook.

Rich, A. (1984). *The fact of a doorframe: Poems selected and new, 1950-1984.* New York: W. W. Norton & Co.

Seidman, I. E. (1991). *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* New York: Teachers College Press.

Shimizu, K. (1995). Japanese college student attitudes towards English teachers: A survey. *The Language Teacher, 19* (10), 5-8.

Terasaki, M. (1995). The position of teachers in modern Japan: Academic freedom, classroom freedom, and job security. *The Language Teacher, 19* (3), 20-22, 25.

Widdows, S., & Voller, P. (1991). PANSI: A survey of the ELT needs of Japanese university students. *Cross Currents, 18* (2), 127-141.

Wordell, C. (1993). Politics and human relations in the Japanese university. In P. Wadden (Ed.), *A Handbook for Teaching English at Japanese Colleges and Universities* (pp. 145-155). Oxford: Oxford University Press.

APPENDIX A

KGU Assessment Survey: Evaluation in English Classes

Please choose the class which you teach most often. When specific examples are called for, please only include those answers which relate to *that* class.

Thank you in advance for your cooperation at this busy time of year!

This survey takes 20~30 minutes to complete.

GENERAL INFORMATION

Directions: Please check the appropriate box in each category.

CATEGORY	CHOICES: Check (✔) one for each category.			
employment	❑ full-time	❑ part-time	❑ full-time (contract basis)	
native language	❑ Japanese	❑ English	❑ other (write in ➔)	
*Please check **only one**. (the one you teach most often)*	❑ Eigo I Ko ❑ IM English ❑ Eigo II ❑ Eikaiwa ❑ Eigo Hyougen Enshu ❑ Eigo D	❑ Eigo II Ko ❑ Pre-Ad English ❑ Eigo III ❑ Eikaiwa I ❑ Eigo A ❑ Ei Sakubun I (Composition I)	❑ Eigo I Otsu ❑ Advanced English ❑ Eigo IV ❑ Eikaiwa II ❑ Eigo B ❑ Ei Sakubun II (Composition II)	❑ Eigo II Otsu ❑ Eigo I ❑ Eigo V ❑ Eigo Hyougen ❑ Eigo C ❑ other: Please write in course title ➔.
range in level between top and bottom students in the class	❑ non-existent (All students are at approximately the same level.)	❑ slight (The majority of students are within the same general level range, and leveling concerns rarely affect the way I think about my teaching.)	❑ moderate (There are some noticeable differences, but these are not so great as to cause serious problems. Sometimes, however, I am concerned about how to teach to all levels.)	❑ extreme (The highest level students are very proficient; the lowest level students can only use English at a basic level. It is quite difficult to adequately address the diverse levels in the class.)
average class size	❑ less than 10	❑ 11~20	❑ 21~30	❑ 31~40 ❑ 41~50 ❑ more than 50

COURSE CONTENT

We would like to know more about what actually occurs in the class.

SKILLS

Please check (✔) the skill(s) that you focus on in the course you listed above. If there are two or more skills, please also underline the main or primary area of focus.

❑ listening ❑ reading ❑ writing ❑ speaking

GOALS

In one or two sentences, please briefly describe what you hope the students will be able to do by the end of this course.

How did you come to understand this goal? (Check as many as apply.)

- ❑ guidance from my department/school
- ❑ talking with other KGU faculty
- ❑ my own ideas of how to teach this class
- ❑ other (Write in ⇒.)

GRADING

Please list the factors you consider in calculating your students' grades. Please include the percentage assigned to each factor.

Example

Factors (write in)	% of grade*
Final Exam	*70%*
Quizzes/Mid-term	*10%*
Participation	*10%*
Homework	*10%*

If there are any other factors which you consider (for example, in cases of borderline grades), but which you do not assign a percentage of the grade, please list those below.

attendance

Your Class

Factors (write in)	% of grade*

If there are any other factors which you consider (for example, in cases of borderline grades), but which you do not assign a percentage of the grade, please list those below ⬇.

*If you do not decide the percentages at the start of the semester, please explain:

_____

Do you inform your students about how you grade them at/near the beginning of the class?

- ❑ I explain which factors I will count toward the final grade as well as what percentage I will assign to each factor.
- ❑ I only explain which factors will be graded.
- ❑ I do not explain how I will grade.
- ❑ other (Write in ⇒)

Do you have complete control about how to grade? ❑ yes ❑ no

If "no," please briefly explain any rules or restrictions which you have been asked to follow in grading your students.

Example: *I was told that the final exam must be at least 50% of the final grade.*

Approximately what percentage of students in your class fail? _____%

DESCRIPTION OF ASSESSMENT CRITERIA

We would like to know how you are using various assessments in your class. Please provide a brief description of each kind of assessment listed below which you are using in your class. Your explanations about why you have chosen certain assessment techniques will be especially helpful to us. **Although this area will be the most time-consuming to fill out, it will also be the most useful**

to us in terms of our research purposes. It will be of greatest help to us if you provide as much detail as possible.

Note: Two sample descriptions have been provided for you.

Example #1
1. Final Exam

Task/Activity: Three Parts:
1) vocabulary translation (into English)
2) a short passage translation (I give them a paragraph of about 300 words and ask them to paraphrase the overall meaning. Usually, the topic is related to current news items in some way.)
3) a dictation (The dictation is made up of about 100 words). All of the test items come from the short quizzes and textbook assignments they have throughout the semester, except I choose a new dictation because I want them to see their progress.

Evaluation:
1) If the student provides the exact same translation as was in the textbook, I give full credit for that item.
2) I give full points to students who can capture the overall ideas adequately while using at least five of the new English expressions they learned and writing in a clear and easy-to-understand way. I take off points if the student has misinterpreted the meaning, not used any expressions studied in class, or miss-spelled words.
3) I give partial credit if the student has the basic word correct but missed the form (for example, she/he wrote "listen" instead of "listened." Also, I only give half credit if the student has miss-spelled the word.

Example #2
6. Informal Assessments (ungraded class activities, etc.)

Task/Activity: I try to do this with all tasks, but especially those that involve group work.

Evaluation: Although I don't grade this factor, I do try to judge what students are ready for by what they do in class. For example, if a student seems really comfortable with speaking in front of the class, I might put her/him in a small-group with other students who need some advice about how to prepare a speech. Or, if I notice that another student doesn't yet seem to understand a grammar point we've already covered, I might give that student an extra exercise on that.

Please only write in answers for those categories you actually consider in your classes. Please be sure to separate the activity and evaluation descriptions when you write your answers. Use the back of page five if you need more space.

1. Final Exam

Task/Activity:

Evaluation:

2. Quizzes, Mid-terms, and/or other Tests

Task/Activity:

Evaluation:

3. Class Participation

Task/Activity:

Evaluation:

4. Homework

Task/Activity:

Evaluation:

5. Projects/Essays

Task/Activity:

Evaluation:

6. Informal Assessments

Task/Activity:

Evaluation:

OVER →

7. other (write in)

Task/Activity:

Evaluation:

FINAL QUESTIONS

Are you satisfied with the way you are assessing your students? Please briefly explain your answer.

What do you think is the biggest problem or challenge that university English teachers in Japan face in terms of assessing student progress?

If you could change anything about your current teaching and/or assessment situation, what would you like to change?

Thank you very much for your valuable time!